

**HIGH QUALITY NUTRITION IN CHILDHOOD AND WAGES IN EARLY
ADULTHOOD: A TWO-STEP QUANTILE REGRESSION APPROACH FROM
GUATEMALAN WORKERS**

MARIA CECILIA CALDERON¹

ABSTRACT

Establishing a causal relationship between health and productivity is not straightforward. On one hand, higher income individuals invest more in health: as their income grows, they invest in better diets and health care. On the other, if a worker is healthier and more energetic, then she will probably be more productive. This paper focuses on the second pathway and examines the effect of one dimension of health, height and body mass index (BMI), on wages. Data comes from a longitudinal study conducted in Guatemala, a low-income country, during 1969-1977 and followed-up in 2002-2004. The estimates suggest a very non-linear relationship between height, BMI and wages; however, the evidence is stronger for males than for females. While diminishing returns are operating at higher quantiles of the conditional wage distribution, increasing returns appear at lower quantiles, implying that height and BMI might have an increasing payoff for the poorer workers.

Key words: Health; Height; BMI; Wages; Guatemala; Quantile.

JEL classification: J24; I12

March 2007

¹ Maria Cecilia Calderon, Population Studies Center, University of Pennsylvania, Philadelphia, PA. E-mail: cmaria@sas.upenn.edu. Paper presented at the Annual Meeting of the Population Association of America, March 29-31, 2007, New York, NY. This paper has benefited from the valuable comments of Jere Behrman, John Hoddinott, Hans-Peter Kohler, John Maluccio, Reynaldo Martorell, Manuel Ramirez, Erica Soler-Hampejsek, and Aryeh Stein. The usual disclaimers apply. This research was conducted with support from the NIH/Fogarty grant TW-05598 on "Early Nutrition, Human Capital and Economic Productivity," NSF/Economics grants SES 0136616 and SES 0211404 on "Collaborative Research: Nutritional Investments in Children, Adult Human Capital and Adult Productivities," NIH grant HD046125 on "Education and Health over the Life Course in Guatemala," R01 HD045627-01 grant on "Resource Flows Among Three Generations in Guatemala," NIH/NIA grant P30 AG12836 to PARC at the University of Pennsylvania, the Boettner Center for Pensions and Retirement Security at the University of Pennsylvania and the Mellon Foundation grant to the Population Studies Center of the University of Pennsylvania.

1. INTRODUCTION

This research attempts to explore the link between nutrition and productivity within the context of a developing country, Guatemala. The aim is to establish the causal effect of adult height and body mass index (BMI), defined as the ratio between weight (in kilograms) and height (in meters) squared, upon current wages using data collected in four poor Guatemalan villages, settings where returns to physical strength and energy may be substantial.

It is intuitively appealing to believe that better nourished individuals are more productive. Furthermore, the structure of employment in lower income economies is such that work often relies more heavily on physical characteristics such as strength and stamina, and therefore, on good health. However, the nutrition-productivity link is complex to establish. Although it is natural to assume that improved nutritional status leads to increased productivity; it is equally plausible that increased productivity leads to higher income which, in turn, improves nutritional status. This feedback between nutrition and productivity suggests that the labor market consequences of poor health are likely to be more serious for the poor who are more likely to suffer from severe health problems and to be working in jobs for which strength has a payoff.

Returns to health (or nutrition), one of the dimensions of human capital, have been widely analyzed from a theoretical perspective in the labor literature; however, there is not much empirical validation on this topic. On the contrary, returns to education, another aspect of human capital, appear substantial in labor markets and a widespread evidence has emerged for almost every country in the world: Arias, Hallock, and Sosa Escudero (2001) for Argentina; Buchisnsky (2001) for the US; García, Hernández and López-Nicolás (2001) for Spain; Machado and Mata (2001) for Portugal; Montenegro (2001) for Chile; and Tannuri-Pianto and Pianto (2002) for Brazil. All of these examples estimate returns to education in Latin American countries using quantile regressions. Also, there are many studies that report estimates of significant returns to schooling on the average (e.g., see Psacharopoulos and Patrinos (2004) for a survey for developing countries).

The relationship between health and market outcomes has been controversial and comparatively much less explored. Although this link has played a key role in theories of economic development, through the idea of efficiency wages, former empirical studies on this subject have typically concluded there is little reliable evidence indicating that health has an important impact on labor productivity. Thomas and Strauss (1997) pointed out that this lack of reliability emerges from two causes. First, the small number of studies on the matter reflects the fact that health indicators have rarely been collected in surveys that contain measures of wages or productivity. And second, there is a non-trivial interpretation of correlations between health and labor outcomes; early studies have paid little or no attention to the direction of causality. Thus, these studies ignored the fact that any component of income, such as wages and labor supply, may affect current behavior which, in turn, affects health through the consumption of a high quality diet, and vice versa.

Moreover, Leibenstein (1957) hypothesizes that, relative to poorly nourished workers, those who consume more calories are more productive, and that at very low levels of intake, better nutrition is associated with increasingly higher productivity. As well, Strauss and Thomas (1998) argue that such non-concavities lie at the heart of the efficiency wage models. Employers have an incentive to raise wages above the minimum supply price of labor excluding those workers in poorest health from the labor market because they are too costly to hire.

Previously, some researchers have been trying to understand the intricate interrelation between health, nutrition and economic productivity dealing with the potential endogeneity issue. Many of the studies in economics have dealt with this bias of simultaneous effects by developing models which predict the nutrition input variables based on exogenous factors such as prices and household demographic variables, in instrumental variable estimates. For instance, Immink and Viteri (1981a, 1981b) find for sugarcane cutters in Guatemala that it was the leisure time that appeared to be most affected by inadequate energy consumption. Men with low energy consumption decreased the energy intensity of their leisure time activities but not the amount of energy expended at work. When the energy intake increased, the men did not increase the supply of units of work but rather become more active in their leisure time. Another example is found in Immink, Viteri and Helms (1982) again for Guatemalan sugarcane cutters.

Additionally, Strauss (1986), using data from Sierra Leone, uses the predicted household energy intake per capita to explain household farm production. The results suggest that household energy consumption was a positive, significant determinant of farm productivity. A similar approach is used by Sahn and Alderman (1988) with data from Sri Lanka. This study employs predicted household energy consumption per capita as the measure of nutritional status and relates it to wage earnings. Surprisingly, household energy per capita appears a significant, positive determinant of men's but not of women's wages. This differential result between men's and women's productivity is a finding in almost all studies linking nutrition to productivity. Both the Strauss (1986) and Sahn and Alderman (1988) analysis are limited to the use of household energy values as the only measure of individual nutritional condition.

Clearly, a measure of individual nutrient consumption and more importantly an indicator of an individual's nutritional status would have strengthened the analysis. Among other measures of individual nutrition, height and BMI have been widely analyzed in previous research. The best-documented fact in observational studies is that taller people tend to enjoy greater success in labor markets. At the micro level, many studies have demonstrated a positive association of height with hourly earnings. Seminal work by Fogel (1994) has documented secular increases in height which parallel economic growth in the historical literature.

Additionally, Deolalikar (1988) explains wage earnings and farm outputs with measures of both individual energy intake and BMI using data from India. The author finds that even though energy intake is not a significant determinant of wages, BMI appears relevant. Also, BMI, but not energy consumption, has a significant, positive effect on farm output. Thus, nutritional status, proxied by BMI, appears as a key determinant factor in shaping labor productivity.

Furthermore, Thomas and Strauss (1992, 1997) examine the nutrition-productivity link using wage earnings of both employees and the self-employed in urban Brazil. They use four indicators of nutrition as explanatory variables: height, BMI, per capita calorie consumption and per capita protein intakes. Their findings indicate that height is a significant determinant of the wages in urban Brazil: taller men and women earn more even after controlling for education and other dimensions of health. However, BMI is a positive and significant predictor of males' but not for females' wages. These authors suggest that BMI is probably correlated with strength since its effects are largest among the least educated men who are more likely to do manual labor and very physical demanding activities. Also, this research suggests that per capita calorie and protein intake are significantly related to wages but the positive effect of calories disappears rapidly indicating that it may only be the very malnourished for whom energy is a limiting factor for wage earnings. Interestingly, after controlling for height and BMI, calorie intake has diminishing returns; but when protein consumption is added to the model, protein intake has an increasingly effect in wages reflecting the impact of an improved quality diet (measured by the fraction of calories from protein sources). The authors conclude that health (through improved nutrition) provides an important return to labor in Brazil. In addition, Strauss and Thomas (1998) conclude that the positive link between height, BMI and wages is also significant in the US: men who are taller and heavier (given height) earn higher wages.

Moreover, Thomas and Frankenberg (2002a) indicate that even though BMI had no effect on earnings, BMI affected the wages of time-rate workers but not piece-rate workers for adult Indonesian males. They argue that health is difficult to observe and employers use the BMI as a marker for health. As well, these authors find that a 1% increase in height was associated with a 5% increase in earnings, suggesting that taller people are probably stronger, an attribute that is probably more highly rewarded in lower-income settings. Also, they argue that height is a proxy for more than just strength and suggest that part of height is influenced by genotype and reflects family backgrounds. Hence, height is largely determined in early childhood and reflects health and human capital investments made by the parents. Therefore, correlation between height and wages will diminish as the model includes other dimensions of human capital: controlling for age and education cuts the elasticity of wages with respect to height in half for Indonesian males.

Recently work by Thomas et al. (2005) provides unambiguous evidence in support of the hypothesis that health has a causal effect on economic prosperity of males during middle and older ages. The research consists on a random assignment design intervention in which Indonesian adults receive a treatment of iron every week for a year. The findings reveals that males who were iron deficient increase their physical and psycho-social health and economic productivity after the treatment. Also, they appear more likely to be working, sleep less, lose less work time to illness and more able to conduct physically arduous activities. Although benefits for women are in the same direction, the effects are more muted.

The evidence reviewed from earlier studies provides mixed results to explain the nutrition-productivity link. In all of them, height is treated as an indicator of long-term nutritional status and appears to be the variable most often associated with productivity. In addition, in most of these studies, height is treated

as an exogenous variable. Furthermore, many of the previous studies that explore the nutrition-productivity link are limited to males. And in those studies where data is separated by gender, the specific impact of nutrition upon economic output differs among sexes. Consequently, the empirical evidence does not suggest a clear answer for causality in this relationship, particularly in low income countries, where attention has been focused on low levels of BMI. However, although obesity is a central concern in some developed countries, certain concerns with obesity are emerging in poor economies.

The aim of this research is to find support for the positive effect of health, measured by height and BMI, upon wages, at least among those with low height and/or low BMI. Consistent with previous literature (Thomas and Strauss, 1997; Strauss and Thomas, 1998), the non-linear pattern between health and wages is likely to emerge in this research. Furthermore, this paper analyzes the nutrition-productivity link within the context of a developing country, Guatemala. The main objective is to establish a causal relationship between adult height and BMI and current wages using information collected in four poor Guatemalan villages, settings where returns to physical strength and energy may be substantial. Data comes from “The Human Capital 2002-04 Study in Guatemala: A follow up to the INCAP² Longitudinal Study 1969-77” in which a cohort of young men and women, who participated as young children in a randomized community trial of nutrition supplementation, were resurveyed 35 years later. In the original study, two randomly selected villages received a nutritional supplement and two other villages received a control drink. The follow-up study conducted during 2002-2004 collects current data from the former participants. This new report is matched with information about the same individuals 35 years earlier.

Formally, this paper focuses upon the following question: Does an increase in quality diet during childhood affect economic productivity in early adulthood? This issue can be split into different sub-questions: (1) How is the link between wages (labor productivity, under certain assumptions), childhood nutrition, height, BMI and education? (2) Are returns to nutrition homogeneous across workers? (3) If not, why is *quantile regression* an appropriate tool to explore the source of this heterogeneity? (4) How can this research exploit the experiment in the four Guatemalan villages during 1969-1977 to deal with the endogeneity bias in the *quantile regression* framework?

The next sections are organized as follows. Section 2 describes the conceptual framework and formally establishes the central purpose of this research. The econometric model is presented in Section 3 and the experimental data are discussed in Section 4. The main body of evidence is presented in Section 5 where structural equations are estimated and also nonparametric relations are presented. Section 6 summarizes the findings and discusses what conclusions can be made from this work. It also explains limitations to the analysis and some possible further extensions. A detailed description of the variables can be found in Section 7.

² Institute of Nutrition for Central America and Panama.

2. CONCEPTUAL FRAMEWORK

The aim of this research is to estimate the impact of health indicators on labor wages for Guatemalan workers. In order to obtain consistent estimates, two aspects should be considered.

First, health is a multidimensional concept. Several variables can measure health, each of them capturing different dimensions and possibly impacting differently on wages. Some studies have used morbidity incidence as an indicator of health status, however, illnesses indicators may not be accurately captured in surveys since they usually come from self-reports. Other studies, such as Thomas and Frankenberg (2002b), use self-reported anthropometric measures. These authors show that males tend to overstate their height and that above age 50 the overstatements increases with age. Apparently, as men shrink with age, they do not update their height. On the contrary, they do not find a significant level of overstatement in female's height. Interestingly, they also show that while men overstate their height, females overstate their weight. Given this fact, the present analysis employs anthropometric indicators, height and BMI, measured using standard methods, thus, they can be considered less subjective.

Secondly, it is particularly relevant to establish a causal relationship between health and wages in this context. Thus, this paper applies an instrumental variable approach to solve the potential endogeneity bias including a set of instruments. These instruments are assumed to be correlated with observed height and BMI but not with unobserved characteristics that affect wages, height and BMI simultaneously.

Formally, consider a typical wage equation, conditional on health and other individual factors:

$$\ln(\text{wage}_i) = f(h_i, x_i, \varepsilon_i) \quad (1)$$

where $\ln(\text{wage}_i)$ denotes the natural logarithm of wage for worker i , h_i stands for a vector of individual health indicators (height and BMI), x_i denotes other individual characteristics (education, age), and ε_i is an unobserved error term. In fact, only wages for those individuals who work in the labor market can be observed; thus, selectivity into labor force, especially for women, can potentially bias the estimates. Thus, the model is estimated separately for males and females in an attempt to determine gender specific parameters and significance.

The vector of health indicators, h_i , captures a dimension of health measured by two anthropometric indicators: height and BMI. These non-subjective variables appear as outputs of the quality of nutrition during childhood, particularly height, and both are treated as endogenous variables.

First, height is a cumulative measure reflecting both investments in nutrition during one's life (mostly as a child) and also probably infectious disease experience. Considerable literature attempts to establish an association between adult height and productivity, in both developed and developing countries. In the context of a developing country, the usual assumption is that adult height represents long-run nutritional status, determined in substantial part during early childhood. Given this assumption, this literature considers height as statistically predetermined, rather than the output of dynamic investments that individuals make in the presence of persistent genetic and other endowments, as Behrman, Hoddinott and Maluccio (2005) argue. The former treatment seems unconvincing in light of the vast evidence on the effect

of persistent unobserved characteristics such as genetic endowments. For instance, Behrman and Rosenzweig (1999, 2002, 2004, 2005); Behrman, Rosenzweig and Taubman (1994, 1996); Pitt, Rosenzweig and Hassan (1990); Rosenzweig and Schultz (1985, 1987); and Rosenzweig and Wolpin (1995) find that these unobserved characteristics are relevant. Thus, if these unobservable factors are correlated with the observed characteristics, returns to height may be biased since they may confound effects arising from both the observed height and the long-run genetic endowments. Thus, one of the contributions of this study is to consider height as endogenous in order to correct the endogeneity bias.

Second, BMI is calculated from a person's weight and height. BMI does not measure body fat directly, however, the US Centers for Disease Control and Prevention (Department of Health and Human Service) suggest that it is a reliable indicator of body fatness for people. Thus, BMI can be considered an alternative for direct measures of body fat since it is an inexpensive and easy-to-perform method of screening for weight categories that may lead to health problems.

Furthermore, BMI is thought to be correlated with physical capacity and extremes values of BMI have been shown to be related to elevated morbidity and mortality (Thomas and Frankenberg, 2002b). Generally, BMI is suitable for recognizing trends within sedentary or overweight individuals because there is a smaller margin for errors. However, BMI categories do not take into account factors such as frame size and muscularity and the categories do not distinguish what proportions of a human body's weight are muscle, fat, bone and cartilage, or water weight. Despite this, BMI categories are generally regarded as a satisfactory tool for measuring whether sedentary individuals are "underweight," "overweight" or "obese." It has been used by the World Health Organization (WHO) as the standard for recording obesity statistics since the early 1980's. WHO gives the following classification by gender according to BMI:

Males

- Underweight: less than 20 (<20)
- Ideal: greater than or equal to 20 but less than 25 (≥ 20 but <25)
- Overweight: greater than or equal to 25 but less than 30 (≥ 25 but <30)
- Obese: greater than or equal to 30 (≥ 30)

Females

- Underweight: less than 18 (<18)
- Ideal: greater than or equal to 18 but less than 25 (≥ 18 but <25)
- Overweight: greater than or equal to 25 but less than 30 (≥ 25 but <30)
- Obese: greater than or equal to 30 (≥ 30)

BMI depends on the net energy intake and so varies through the life course. It captures both longer and shorter term dimensions of nutrition and it is related to aerobic capacity and endurance (independent of energy intake). Whether this pathway is one through which health importantly influences productivity is not obvious since many jobs do not require sustained physical effort. Some tests suggest that excess weight (fat) affects the efficiency with which energy is transferred to work output.

On the one hand, current BMI may be affected by contemporaneous movements in income or prices; on the other hand, current BMI partly reflects previous health investments. In other words, while better health may result in a worker being more productive, higher income may be spent on improving one's health: this bidirectional relation, or reverse causality, is the key aspect in this study. Thus, the potential correlation between health indicators and the unobserved error term cannot be ignored to obtain consistent estimations.

In addition, equation (1) includes the vector x_i which controls for educational attainment and year of birth. Even though education is not a dimension of health, it is, as well as health, a human capital investment and it is highly related to well-being since it can be seen as a channel through which early nutrition affects current wages. Furthermore, education is likely to be correlated with unobserved individual ability, included in the disturbance term, in the sense that higher ability individuals are likely to acquire more education due to lower implicit costs. Then, this unobserved heterogeneity bias may generate inconsistent estimations in the returns to education. Thus, not only health, but also education is considered endogenous in equation (1).

Following Thomas and Strauss (1997), the implementation of an instrumental variables estimator isolate the effects of health (and education) on wages. Even though these authors treat BMI as endogenous, they consider height and education as exogenous, and hence, they do not instrument these variables. However, the arguments exposed above suggest education is a potential endogenous variable (through the correlation with unobserved ability) and this is the approach used in this empirical work.

3. METHODS

3.1 Consistent estimators

The estimation strategy will be to estimate the effect height and BMI on wages in a quantile regression framework, obtaining unbiased estimates at different quantiles of the conditional wage distribution. The method proceeds as follows.

On one hand, the unobserved error term in (1) may include genetic endowments that can neither be observed nor measured. If these genetic endowments simultaneously affect health and wages, then any OLS estimator of the impact of height and BMI on productivity would be inconsistent. On the other hand, education may be correlated with ability in the sense that more capable individuals acquire more education because the relative cost of schooling is lower. Note that ability is also correlated with wages, thus, education is also an endogenous variable.

This research assumes that health (and education) and genetic endowments (also ability) interact in a non-trivial, unknown way. On one hand, if genetic endowments and health are substitutes in the generation of human capital, then marginal returns to the accumulation of human capital might be expected to decrease with endowments and hence, health contributes relative more to low endowed individuals. In this case, an estimate that ignores the endogeneity bias would be underestimated. On the other hand, if

endowments and health are complements in the generation of human capital, then health has an additional indirect effect on human capital (through the interaction with endowments) that increases its otherwise constant contribution to earnings. In this case, returns would then be higher for the better endowed and any estimate that ignores the endogeneity bias would be overestimated.

A priori, the interaction between health and endowments is unknown and this paper proceeds without observing genetic endowments. Hence, it is not possible to model the relationship between endowments and height and BMI explicitly including additional regressors. Moreover, the key assumption is that health is not randomly assigned to individuals and thus, the current level of BMI or height can not be treated as predetermined. Therefore, the level of health is likely determined endogenously as a function of the level and quality of genetic endowments and other factors such as family background and individual and community characteristics.

Consequently, all these exposed arguments suggest the use of an instrumental variable (IV) estimator for both height and BMI with the purpose of isolating the causal effect on productivity. Variables correlated with health (education) but, at the same time, uncorrelated with the unobserved disturbance in the wage relation (1) can serve as valid instruments. Thus, the instruments or exogenous variables employed in the analysis come from the original study conducted during 1969-1977 and correspond to background and family characteristics, such as the socioeconomic status and parental education, as well as individual and community level characteristics, like the potential exposure to the supplemental drink during the first years of life (since current height and BMI are associated with the quality of nutrition during childhood), quality of the school and the ratio of students to teachers when the individuals were seven years old. The underlying assumption is that the instruments have a direct effect upon height and BMI but do not affect current wages except through their impact upon those health indicators. Such assumption seems realistic considering that employers may not directly observe the nutritional status of the workers but instead observe their current height and BMI (which are the result of past nutritional investments) and pay wages according to them. The key point is that a favorable background at family, individual and community level is correlated with current height and BMI. In other words, this model assumes that individuals who were exposed to the nutritional supplement in the first two years of life, those who lived in the experimental villages, are currently better nourished, and this attribute is reflected in height and BMI.

3.2 Heterogeneity in marginal effects

A mean regression, whether instrumented or not, is the typical analysis in the economic and health literature. Thus, this usual method only estimates the effects of health on average wages. These models are restrictive because they omit the effects at other parts of the conditional (on the observed characteristics) distribution of wages. Contrary to a mean regression, this paper aims to estimate the impacts of height and BMI at different quantiles of the conditional distribution of earnings. For example, the impact of height at the 10% conditional quantile may be significantly different from the impact of height at the 90% conditional quantile, implying heterogeneity in the returns to height. For this reason, a quantile regression

approach (Koenker and Basset, 1978; Koenker and Hallock, 2001) may provide very interesting results when estimating the relationship between height, BMI and productivity. Furthermore, very non-linear relationships between health and wages are expected, and the non-linearities might differ at different quantiles. Consequently, in order to obtain a widespread picture of how height and BMI impact different quantiles of the conditional wage distribution, model (1) is estimated using a quantile regression (QR) approach. See Section 3.3.1 for a formal description of this technique.

3.3 Combining Instrumental variables (IV) and Quantile regression (QR)

Two-step quantile regression estimates yield a family of quantile estimators while simultaneously correct the endogeneity bias. Amemiya (1982) first proposed this method, followed by Powell (1983) and Chen and Portnoy (1996), who extended the first established properties of this consistent estimator.

The present research pursues two main goals: (1) to give a widespread picture of the effect of health on productivity over the entire conditional wage distribution, not only on the mean; and (2) to obtain consistent estimates of such impacts. Thus, this paper combines both purposes through the use of two-step quantile regression; further details are given in Section 3. For instance, Arias et al. (2001) apply this technique to consistently estimate returns to education.

Thus, two-step quantile regression (IV-QR) isolates the effect of height and BMI on wages and produces unbiased estimators at different quantiles of the conditional wage distribution. Moreover, different non-linearities, diminishing or increasing returns to health, are expected at different quantiles of the conditional wage distribution. See Section 3.3.2 for a formal description of this technique.

To sum up, the contribution to the literature in this research include: (1) the adoption of a quantile regression approach in order to account for heterogeneity in the returns to health and schooling at different quantiles of the conditional wage distribution; (2) the treatment of education as an endogenous variable; and (3) the treatment of height as an endogenous variable. Even though some studies consider education as endogenous; however, they do not include health indicators in the wage equation. On the other hand, when health measures are taken into account, only BMI is typically treated as endogenous and height and education appear as predetermined. Thus, this research improves previous work treating simultaneously height, BMI and schooling as endogenous variables.

3.3.1 Quantile regression

A quantile estimation can be defined as in Koenker and Basset (1978) and Koenker and Hallock (2001) as the solution to the problem of minimizing a weighted sum of absolute residuals. The τ -quantile in a sample of n observations $\{y_1, \dots, y_n\}$ can be computed by

$$\min_{\xi \in \mathfrak{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - \xi) = \sum_{i=1}^n [\tau I(y_i > \xi) + (1 - \tau) I(y_i < \xi)] |y_i - \xi| \quad (2)$$

where I denotes the indicator function that takes a value of one if the event is true and zero otherwise. Consequently, the conditional linear quantile of Y is estimated by replacing ξ by $x_i'\beta$, where β are the coefficients for the τ^{th} quantile and x_i is a matrix of p explanatory variables, and solve

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - x_i' \beta) \quad (3)$$

The resulting regression fit $x_i'\beta$ describes the τ th quantile of the response variable y_i (the wage of worker i) given the vector of characteristics x_i (height, BMI, years of schooling, age). Thus, the solution to the linear problem above yields a vector of p estimated coefficients for each quantile τ , and β can be seen as $\beta(\tau)$. The full sample of n observations is used in the estimation of each quantile and there is no loss in estimating as many quantiles as desired. Consequently, quantile regression is more general than a simple mean regression, and is extremely powerful when the $\beta(\tau)$ coefficients differ significantly across quantiles, suggesting that the marginal effect of a particular variable, returns to nutrition in this paper, is not homogeneous across τ 's.

3.3.2 Two-step quantile regression

Quantile regressions on a wage equation like (1) yield inconsistent estimates of the returns to nutrition (and education) in the same way OLS delivers an inconsistent estimate of the mean return if predictor variables are correlated with the unobserved error term. The previous section recognizes the existence of unobservable factors (genetic endowments, ability and unobservable background characteristics and family effects) that will be correlated with the observed regressors (height, BMI, education), making the causal interpretation difficult, as pointed out in Behrman, Hoddinott and Maluccio (2005) and Thomas and Strauss (1997). This endogeneity bias can be solved adopting an instrumental variable approach in the quantile regression framework. This mixed method, called two-step quantile regression, combines both techniques. As already mentioned, the first one in proposing this routine was Amemiya (1982)³.

Consider the following structural model:

$$Y = X_1\gamma + X_2\beta + \varepsilon \quad (4)$$

where Y is the response variable, X_1 is a matrix of k_1 endogenous variables correlated with the error term ε (like health and education), X_2 is a matrix of k_2 exogenous regressors (like age), and γ and β are vectors of associated coefficients respectively.

Collecting a set of z instruments in the matrix Z , quantile regression is combined with the classical instrumental variable approach to consistently estimate heterogeneity across quantiles of the conditional wage distribution. The method proceeds in two steps. The first stage projects each endogenous variable contained in X_1 on the space spanned by the instruments (included in Z) and by the exogenous variables (included in X_2), which are, by assumption, uncorrelated with the error term. Thus, the first step is a typical

³ Amemiya called the technique 'Two Stage Least Absolute Deviations Estimators' (2SLAD).

OLS regression of the endogenous variables on the instruments. The second stage performs quantile regressions of the dependent variable on the fitted values from the first step, \hat{X}_1 , and on the exogenous variables, X_2 .

The reduced form equations for Y and X_1 corresponding to model (4) are as follows:

$$Y = X \Pi + V \quad (5)$$

$$X_1 = X \Pi_1 + v \quad (6)$$

where $X = [X_2, Z]$ is a $n \times (k_2+z)$ matrix grouping all the exogenous variables, and V and v are independent and identically distributed error terms. The reduced form equation (5) gives an estimate the effects of the instruments (and the exogenous variables in X_2) on the response variable Y . The reduced form equation (6) shows the effect of the exogenous variables (X_2 and Z) on the endogenous variables (X_1). The asymptotic properties of this two-step quantile regression estimator were proved by Powell (1983), Chen (1988), and Chen and Portnoy (1996).

In this framework, equation (5) represents the relationship between individual, family and community backgrounds (and all other exogenous variables) and wages. Analogously, equation (6) represents the effect of individual, family and community backgrounds on health indicators (height and BMI).

4. DATA

This paper uses a longitudinal data set collected over a 35 year period in four poor Guatemalan villages (San Juan, Conacaste, Santo Domingo and Espiritu Santo) by the Institute of Nutrition for Central America and Panama (INCAP). A more complete report and further details can be found in see Grajeda et al. (2005), Hoddinott, Behrman and Martorell (2005), Maluccio, et al. (2005b), Martorell et al. (2005) and Stein et al. (2005).

During 2002-2004, a team of researchers undertook a follow-up data collection on the participants in a randomized trial intervention during the period 1969-1977. The original INCAP Longitudinal Study was recorded for children 7 years or younger, so the year of birth for the participants ranges from 1962 to 1977, implying that these participants were 0 to 15 years old. The length and timing of exposure to the nutritional interventions for particular children depended on their respective birth dates. For example, only children born after mid-1968 and before October 1974 were exposed to the nutritional intervention for all of the time they were from six to 36 months of age, which often is posited to be a critical time period for child growth in the nutrition literature.

By the time of the 2002-2004 data collection, sample members ranged from 25 to 42 years of age; from the original sample of 2,393 individuals in 1969-1977, approximately 4% were untraceable, 11% had died and 8% had migrated abroad. This fact may lead to systematic bias that may invalidate the estimates

due to attrition. However, Maluccio et al. (2005a), using the same data source, find that there is no attrition bias in this sample.

The principal hypothesis underlying the 1969-1977 intervention was that improved pre-school nutrition accelerates physical growth and mental development. To test this hypothesis, 300 villages were screened to identify those of appropriate size, compactness, ethnicity, diet, educational levels, demographic characteristics, and nutritional status. From this screening, village pairs similar in these characteristics were determined: Conacaste and Santo Domingo (relatively crowded villages) and San Juan and Espíritu Santo (relatively less crowded villages).

Two villages, Conacaste and San Juan, were randomly assigned to receive a high protein-energy drink, Atole, as a nutritional supplement. Atole contained Incaparina (a vegetable protein mixture developed by the INCAP), dry skim milk, and sugar and had 163 kcal and 11.5 g of protein per 180 ml cup. This design reflected the prevailing view of the 1960's that protein was the critically limiting nutrient in most developing countries. Atole, the Guatemalan name for hot maize gruel, was served hot; it was pale gray-green and slightly gritty, but with a sweet taste.

In designing the data collection, there was considerable concern that the social stimulation associated with attending feeding centers, such as the observation of children's nutritional status, and the monitoring of their intakes of Atole, also might affect child nutritional outcomes, thus confounding efforts to understand the impact of the supplement. To address this issue, an alternative drink, Fresco, was provided in the remaining villages, Santo Domingo and Espíritu Santo. Fresco was a cool, clear-colored, fruit-flavored drink. It contained no protein and only sufficient sugar and flavoring agents. It contained fewer calories per cup (59 kcal/180 ml) than Atole. Several micronutrients were added to the Atole and Fresco in amounts that achieved equal concentrations per unit volume. This was done to sharpen the contrast between the drinks to protein; the energy content differed, of course, but this was not recognized to be of importance at the time.

The nutritional drinks (Atole or Fresco) were distributed in supplementation centers and were available daily, on a voluntary basis, to all members of the community during times that were convenient to mothers and children but that did not interfere with usual meal times. Interestingly, Schroeder, Kaplowitz and Martorell (1992) show a large differential in the nutritional intake between Atole and Fresco villages. Averaging over all children in the Atole villages (i.e., both those that consumed any supplement and those who never consumed any), children 0-12 months consumed approximately 40-60 kcal per day, children 12-24 months consumed 60-100 kcal daily and children 24-36 months consumed 100-120 kcal per day as supplement. In contrast, children in the Fresco villages consumed virtually no Fresco between the ages of 0-24 months (averaging at most 20kcal per day) with this figure rising to approximately 30 kcal daily by age 36 months. Micronutrient intakes from the supplements were also larger for Atole than Fresco villages; also, the Atole contributed significant amounts of high-quality protein, while the Fresco contributed none.

Given this large differential exposure to treatment, this study exploits the intensive structure of the longitudinal survey and observational work associated with the intervention to construct the variables used

as instruments. The key point is that these instruments, which capture the exposure to the nutritional supplement, are correlated with current height and BMI but not with wages.

This paper includes 680 wage earners⁴ (original participants with a 66% of males) resurveyed in 2002-2004 for whom the two measures of health central to this analysis, height and BMI, are both available. The wage production function (1) includes human capital variables, height, BMI and education, which are all treated as endogenous. Thus, data requirements are substantial in order to construct instruments and get consistent returns to health estimates. While the original study conducted during 1969-1977 provides information to construct the instruments, measures of wages, health and education come from the 2002-2004 follow-up. Section 7 gives a full description of all variable definitions from both studies.

Response variable: the dependent variable in the earning production function (1) is the hourly income from wages. Figure 1 shows the distributions of the logarithm of wages for four groups: males with no schooling, males with some years of schooling, females with no schooling and females with some years of schooling. As is typical for income distributions, the wage distribution is closer to log-normal than to normal. As expected, the wage distribution for males with some schooling is to the right compared to the other categories, implying that educated men have, on average, higher earnings. Additionally, Table 1 presents descriptive statistics for the main variables. Consistent with Figure 1, males have a larger mean hourly wage than females, 10.9 and 7.9 quetzals⁵, respectively.

Key independent variables: three key variables are included as endogenous. First, height was measured to the nearest 0.1 cm, with the subjects bare footed, standing with their backs to a stadiometer. Second, weight was measured on subjects dressed in their normal underclothes with no shoes or objects in their pockets. This measure was taken using a digital scale with a precision of 100 grams. Then, BMI becomes the ratio between weight (in kilograms) and height (in meters) squared. Finally, completed years of formal schooling (adult and informal education are excluded) is the indicator of educational attainment.

Furthermore, Table 1 shows that while the average male is taller than the average female (162.8 and 150.6 cm, respectively), women show, on average, higher BMI than men (26.4 and 24.7, respectively). Additionally, there is, as expected, a positive and statistically significant at 1% correlation between wages and height for both males and females (approximately 0.226); see Table 2. However, wages and BMI are positively correlated for males (0.18, significant at 1%) but not for females (-0.007, not significant). Also, there is a positive and significant at 1% correlation between wages and schooling attainment for both sexes (0.40 for males and 0.34 for females).

In addition, schooling is positively correlated with adult height (0.22 for males and 0.24 for females, both statistically significant at 1%). Surprisingly, there is a positive correlation between schooling

⁴ Two outliers with hourly wage greater than 90 quetzals (local currency) were dropped from the sample.

⁵ At the time of the survey, a dollar was equivalent to 7.59 quetzals.

and BMI for males (0.11), but negative for females (-0.15), both statistically significant at 5%. Finally, there is a positive correlation between height and BMI for men (0.03) but a negative one for women (-0.032). However, both are very weak and not statistically significant.

Finally, Figures 2, 3 and 4 illustrate the distribution of height, BMI and schooling, respectively. Consistent with the tables, men are, on average, taller than women; Figure 2 also shows that males (females) with some schooling are taller than males (females) with no schooling, verifying the positive correlation between height and schooling. Figure 3 reveals that the mean BMI is higher for women than for men; however, the picture does not evidence a clear association between BMI and schooling. So, how is schooling distributed? Figure 4 shows an unequal distribution by gender with a mean value of approximately 5 and 4.5 years for males and females, respectively. The highest frequency is at six years (31% for males and 22% for females), were primary school is completed. There are secondary modes at zero grades (15% for males and 18% for females) and three grades (8% for males and 12% for females).

Instruments: the instruments come from the original study conducted in 1969-1977 and are grouped in three categories:

(1) Individual level: includes four binary variables that equal one if the individual was exposed to the program for when 0-24 months; if the individual was exposed to the program for when 0-24 months and lived in one of the two Atole villages; if the individual was exposed to the program for when 24-42 months; and if the individual was exposed to the program for when 24-42 months and lived in one of the two Atole villages.

(2) Family level: includes a binary variable that equal one if the individual is a twin, Socio Economic Status, dummy equals one if mother died before the subject was age 12, dummy equals one if father died before the subject was age 12, mother's total completed grades of schooling, and father's total completed grades of schooling.

(3) Community level: includes three binary variables that equal one if the child lived in one of the two Atole villages (Conacaste and San Juan); if the child lived in a large village (Conacaste and Santo Domingo); if the child lived in large village and received Atole (Conacaste). This group also includes the number of primary students per primary teacher when the individual was seven years, a dummy equals one if preschool/kindergarten was available in the community when the individual was two years, a dummy equals one if primary school housed in a permanent/modern structure when the individual was seven years, dummy equals one if community receiving regular intrahousehold water service when the individual was two years.

As explained, the main assumption is that these three groups of variables are correlated with current health indicators, height and BMI (also with schooling). However, there is no reason to expect that these three groups of variables are correlated with current wages other than indirectly through schooling and health indicators.

5. RESULTS

5.1 Parametric approach

The main focus of this section is to estimate returns to nutrition across quantiles of the conditional wage distribution after accounting for the endogeneity issue. In other words, the aim is to determine: (1) Do health indicators have a true impact on productivity?; (2) If yes, how large is that impact?; (3) Who experiences the impact?; (4) Is this effect homogeneous or heterogeneous across quantiles?

The preferred model includes height, height squared, BMI, BMI squared, completed years of formal schooling and year of birth⁶ as covariates. Squared terms are incorporated to capture non-linearities. Height (and its squared), BMI (and its squared) and schooling are treated as endogenous variables. Four alternative specifications were tested: (1) only height as health indicator; (2) height and height squared; (3) height and BMI; (4) height treated as exogenous. For comparison, the preferred model, as well as the alternatives, was estimated by OLS, IV and QR. The estimates from this sensitivity analysis are not shown in this paper but are available from the author upon request. The main result is that the non-instrumented estimates are downward biased.

Table 3 shows the estimates from the two-step quantile regressions where height and BMI (and schooling) are treated as endogenous, thus, jointly determined with wages. The coefficients for the wage equation (1) are reported separately for males and females. The method, fully detailed in Section 3, consistently estimates the effects of health (or nutrition) at different quantiles of the conditional wage distribution. The previous section describes the instruments used in the first step, for further definitions see Section 7. Standard errors, based on the bootstrap method with 5,000 replications, are showed in parentheses.

The regression results should be interpreted in the context of Guatemalan villages, poor and rural areas in a developing country. The results imply that the effect of health on wages is clearer for males than females, likely because of the structure of the labor market in those settlements. One limitation of the analysis is the selectivity bias that may potentially arise for females. This fact should be taken in mind.

Consistent with previous work, the estimates for men appear more robust than for females. Even though increasing returns appear at lower quantiles, diminishing returns emerge at higher quantiles of the conditional wage distribution, and this result is similar for both BMI and height.

The estimates show a negative sign for height and positive for height squared at 1% and 5% quantiles (increasing returns). However, note that they are not statistically significant. On the contrary, above the 10% quantile height shows diminishing returns (positive coefficient for height and negative for height squared) and most of the quantiles are statistically significant at 10%. This implies that for men with mean height (163 cm) a 1 cm increase raises wages in approximately 13.8% at the 25% quantile. In addition, a male with average height may expect a 9% increase in wages from a 1 cm increase in height at the 75% quantile. Figure 5 graphically shows the results for the two extreme quantiles, 5% and 95%, and

⁶ Year of birth is reported more accurately than age in the survey.

the median, 50%. The lowest quantile presents a slightly convex pattern between height and wages and the relation becomes concave at higher quantiles.

In the case of BMI, there is also increasing returns at lower quantiles and diminishing returns at higher quantiles. Although some coefficients are not statistically significant under the conventional normal method, they appear significant under the bias corrected confidence interval. Interestingly, for those with an average BMI (25) a one unit increase reduces wages in 8% for those at the 5% quantile. In contrast, for those with a higher BMI (28) a one unit increase raises wages in 3% at the same 5% quantile. Furthermore, for those who are high in the conditional distribution of wages, for example at the 99% quantile, a one unit increase in BMI reduces wages in 5% for the average BMI; however, a lower BMI, such as 24, raises wages in 12%. Figure 6 displays the different non-linearities at some extreme quantiles (5% and 95%) and the median.

Moreover, Table 4 presents tests of equality between coefficients at different quantiles. The null hypothesis is equality of returns. A low p-value rejects the null implying statistically significant differences. For example, the coefficient for height at the 5% quantile is statistically different from the coefficient at 25% (p-value 0.006).

Consistent with prior evidence, returns to health for women are not statistically significant in almost all quantiles. Only at extremes quantiles, 10% and 99%, BMI appears significant. For expositional purposes, Figures 7 and 8 show the estimated relation between height and wages (Figure 7) and BMI and wages (Figure 8) at different quantiles, 5%, 50% and 95%. Compared to the previous figures for men, the effects of health on women earnings are more linear. However, there is a key difference compared with men: schooling appears significant for females in most of the quantiles, which is not the case for males. This suggests that labor markets are differently structured for men and women.

5.2 Nonparametric approach

This section presents evidence of the association between height, BMI and wages using locally weighted regressions. This method is extremely useful to capture non-linearities since it uses locally weighted linear regression to smooth the data⁷. First, this section presents bivariate nonparametric estimates for males. Subsequently, Section 5.2.1 presents some multivariate results for both males and females.

For visual analysis, Figures 9 and 10 show nonparametric regressions of the association between height, BMI and wages (in logarithm scale) for males. Bootstrapped 95% confidence intervals (100 replications) suggest that the nonparametric estimates are statistically significant in all the cases.

For instance, Figure 9 presents a nonparametric regression between height and wages by schooling. The smoothed curves are somewhat linear and almost parallel when comparing males with no schooling and those with some schooling. One may argue that height can be considered as a proxy for education; thus, some of the association between height and wages might be attributed to education.

⁷ Although Stata 'lowess' command selects a default bandwidth of 0.8, I use a bandwidth of 0.85 to avoid undersmoothing the results.

However, as Figure 9 reveals, the correlation persists for those males with no schooling at all. This finding is consistent with the pattern found by Strauss and Thomas (1998) for male Brazilian workers.

In addition, Figure 10 shows the association between BMI and wages for men (by schooling). Interestingly, the relationship between BMI and wages is non-linear. As expected, the smoothed curve for males with some schooling is above the curve for those with no schooling. While both smoothed curves have a positive slope for intermediate values of BMI (between 20 and 30), an increase in BMI is associated with higher returns for males with no schooling. It is plausible that for those men, elevated BMI is associated with greater physical strength, which is of value for manual labor, but that strength is of less value among the better educated who might be more likely to have sedentary occupations. Therefore, the positive correlation between wages and BMI persists for those with no schooling; however, it becomes negative when BMI exceeds 30. Thus, extreme values of BMI, which correspond to higher risk of diseases, show lower outcomes.

5.2.1 Multivariate nonparametric approach

For expositional purposes and visual analysis, this section presents multivariate nonparametric estimates for the association between health and wages. The pictures are presented for males and females separately. Although they are interesting, it is important to remember that they do not control for more than two variables at a time and they do not control for the endogeneity of human capital stocks. Thus, wages are nonparametrically estimated as a function of: (1) BMI and schooling, but not height (Figures 11 for males and 14 for females); (2) height and schooling, but not BMI (Figures 12 for males and 15 for females); and (3) height and BMI, but not schooling (Figures 13 for males and 16 for females).

Therefore, the nonparametric results confirm that the relations among height, BMI and wages are very non-linear. In the case of males, the higher wages are for those better educated and with a moderate BMI (Figure 11). When controlling for height in Figure 12 (but not for BMI), schooling appears relevant, however, being taller is also associated with higher wages even with a medium level of schooling. Moreover, Figure 13 which does not control for education suggests that taller men with an intermediate BMI are richer.

Conversely, results for females are different. Figure 14 shows that a higher level of schooling and a lower BMI are associated with higher wages. When controlling for height in Figure 15 (but not for BMI), the relation is pretty flat, however, there is an optimal combination of height and schooling that yields a maximum level of wage: high schooling and medium height. Figure 16, which excludes schooling, reveals that increasing height and reducing BMI might be associated with higher earnings. However, for the shortest females, a large BMI also is associated with higher wages.

6. CONCLUSIONS AND FURTHER EXTENSIONS

Establishing a relationship between health and productivity is not straightforward. It is likely that causality runs in both directions. On one hand, higher income individuals invest more in human capital, including health: as their income grows, they invest in better diets, improved sanitation and better health care. On the other, if a worker is healthier, less susceptible to disease, and more alert and more energetic, then he or she will probably be more productive and experience higher earnings. This paper focuses on the second pathway and examines the effect of a dimension of health, measured by height and BMI, on wages, an indicator of labor productivity. Data come from a longitudinal study originally conducted during 1969-1977 and followed-up during 2002-2004 in Guatemala, a developing country.

This research consistently estimates returns to height and BMI at different quantiles of the conditional wages distribution using a two-step quantile regression approach. This method provides a widespread picture of health effects, rather than a mean effect. Consistent with previous research, the evidence is more robust for males than for females, suggesting that labor markets are differently structured for men and women.

Moreover, there is a very non-linear relationship between height, BMI and wages. The findings imply that even though diminishing returns to height and BMI (concave shape) are operating at higher quantiles of the conditional wage distribution, increasing returns (convex shape) appear at lower quantiles. Thus, height and BMI might have an increasing payoff for the poorer workers (those that are low in the conditional wage distribution). In addition, schooling appears significant for females in most of the quantiles, which is not the case for males. Finally, nonparametric estimates verify the presence of non-linearities.

Further extensions to the analysis comprise:

- include alternative measures of productivity as outcome variable, such as hours worked per week, which may include hours of housework.
- include alternative measures of body composition to capture physical strength and energy such as skinfold thicknesses and circumferences (Ramirez-Zea et al. 2006).
- include alternative instruments. For instance, food prices might serve as instruments for height and BMI (and schooling attainment), as in Thomas and Strauss (1997).
- include interaction terms between schooling and height and between schooling and BMI.
- account for the selectivity bias into the labor markets including hazard rates, as in Heckman (1974).

7. VARIABLE DEFINITIONS

7.1 Response variable

lwage Hourly income from wages (logarithm scale). The mean hourly wage is higher for men (10.9 quetzals) compared to women (7.9 quetzals). See Figure 1.

7.2 Endogenous variables

height Measured to the nearest 0.1 cm, with the subjects bare footed, standing with their backs to a stadiometer (GPM, Switzerland). All measurements were done twice. If the difference between the two first measurements was greater than 1.0 cm for height, a third measurement was done and the two closest measurements were used. The sample used in this research yields an average value of 162.8 cm for males and 150.6 cm for females. See Figure 2.

BMI Defined as weight (kg) / height squared (m), BMI is a measure of weight for height. It has been promoted as a useful indicator for chronic energy deficiency, and to a lesser extent to indicate obesity. Weight was measured on subjects dressed in their normal underclothes with no shoes or objects in their pockets. The measure was taken using a digital scale with a precision of 100 grams. While men are on average taller than women, females present, on average, a higher BMI compared to men (24.7 for males and 26.4 for females). See Figure 3.

educ Years of completed formal schooling (excludes informal or adult education). The mean is approximately 5 and 4.5 years for males and females, respectively. The highest frequency is at six years, where primary school is completed. There are secondary modes at zero grades and three grades. See Figure 4.

7.3 Controls and instruments

7.3.1. Individual-level variables

byr Year of birth. The INCAP Longitudinal Study was conducted from 1969-1977 and recorded for children 7 years or younger, so the year of birth ranges from 1962-1977.

exp00_24	Dummy = 1 if the individual was exposed to the program for when 0-24 months. It is a cohort measure.
exp00_24a	Dummy = 1 if the individual was exposed to the program for when 0-24 months and lived in one of the two Atole villages ($\text{exp00_24} * \text{atole}$). It is a measure of the Atole intervention.
exp24_42	Dummy = 1 if the individual was exposed to the program for when 24-42 months. It is a cohort measure.
exp24_42a	Dummy = 1 if the individual was exposed to the program for when 24-42 months and lived in one of the two Atole villages ($\text{exp24_42} * \text{atole}$). It is a measure of the Atole intervention.

7.3.2. Family-level variables

twins	Dummy = 1 if the individual is a twin (with zero if missing).
SES	Socio Economic Status. It is constructed as a linear wealth index that is the first principal component of assets owned by the parental household in 1975. For further details see Maluccio, Murphy and Yount (2005).
dthma_12	Dummy = 1 if mother died before the subject was age 12.
dthpa_12	Dummy = 1 if father died before the subject was age 12.
momeduc	Mother's total years of schooling (estimated if the observation is missing).
dadeduc	Father's total years of schooling (estimated if the observation is missing).
dummomeduc	Dummy = 1 if mother's education is missing.
dumdadeduc	Dummy = 1 if father's education is missing.

7.3.3. Community-level variables

atole	Dummy = 1 if community received atole (Conacaste and San Juan). Indicator of whether or not the child lived in one of the two Atole villages.
largevill	Dummy = 1 if the village is relatively large (Conacaste and Santo Domingo).
largevilla	Dummy = 1 if atole = 1 and largevill = 1 (Conacaste).
stu_tch_07	Number of primary students per primary teacher when the individual was seven years. There are gaps in the information provided from school records. To fill these gaps, information from one year is assumed to apply also to the surrounding years.
presch_02	Dummy = 1 if preschool/kindergarten was available in the community when the individual was two years. This variable actually combines two types of preschools. One type is the preschool offered to 3-4 year olds and sponsored by an international organization, the Institute for Cultural Affairs (ICA). The other type is the public preschool sponsored by the Guatemalan Ministry of Education (MINEDUC), and it is offered to 5-6 year olds.
pstruc_07	Dummy = 1 if the primary school was housed in a permanent/modern structure when the individual was seven years (i.e. not wattle and daub, nor adobe, but usually of cement block or bricks). In the year when permanent/modern school structure was completely destroyed by earthquake, this dummy goes to 0 until a new structure was built (Espiritu Santo – Conacaste school was not yet modern).
water_02	Dummy = 1 if the community received regular intrahousehold water service when the individual was two years. Note that in San Juan and Conacaste, water systems were regular at one point and then deteriorated.

REFERENCES

- Arias, Omar, Kevin Hallock and Walter Sosa Escudero, 2001, "Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data", in Fitzenberger B, R. Koenker and J. A. Machado, eds., *Economics Applications of Quantile Regression*.
- Behrman, Jere R., 1993, "The Economic Rationale for Investing in Nutrition in Developing Countries", *World Development*, 21(11): 1749-72.
- Behrman, Jere R. and Anil B. Deolalikar, 1988, "Health and Nutrition", in Hollis B. Chenery and T.N. Srinivasan, eds., *Handbook on Economic Development*, Vol. 1, 631-711, North Holland Publishing Co., Amsterdam.
- Behrman, Jere R. and Anil B. Deolalikar, 1989, "Wages and Labor Supply in Rural India: The Role of Health, Nutrition and Seasonality", in David Sahn, ed., *Causes and Implications of Seasonal Variability in Household Food Security*, 107-118, The Johns Hopkins University Press, Baltimore, MD.
- Behrman, Jere R., Andrew Foster and Mark M. Rosenzweig, 1997, "The Dynamics of Agricultural Production and the Calorie-Income Relationship: Evidence from Pakistan", *Journal of Econometrics*, 77(1): 187-207.
- Behrman, Jere R., John Hoddinott, John Maluccio, Erica Soler-Hampejsek, Emily L. Behrman, Reynaldo Martorell, Agnes Quisumbing, Manuel Ramirez and Aryeh D. Stein, 2005, "What Determines Adult Skills? Impacts of Pre-School, School-Years and Post-School Experiences in Guatemala," Philadelphia, PA: University of Pennsylvania, mimeo.
- Behrman, Jere R. and Mark R. Rosenzweig, 1999, "Ability Biases in Schooling Returns and Twins: A Test and New Estimates," *Economics of Education Review* 18, 159-167.
- Behrman, Jere R. and Mark R. Rosenzweig, 2002, "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?", *American Economic Review* 92(1): 323-334.
- Behrman, Jere R. and Mark R. Rosenzweig, 2004, "Returns to Birthweight," *Review of Economics and Statistics* 86:2 (May), 586-601.
- Behrman, Jere R. and Mark R. Rosenzweig, 2005, "Does Increasing Women's Schooling Raise the Schooling of the Next Generation? – Reply" *American Economic Review* 95 (December).
- Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman, 1994, "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment," *Journal of Political Economy* 102:6 (December), 1131-1174.
- Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman, 1996, "College Choice and Wages: Estimates Using Data on Female Twins," *Review of Economics and Statistics* 73:4 (November), 672-685.
- Buchinsky, Moshe, 2001, "Quantile regression with sample selection: Estimating women's return to education in the U.S.", in Fitzenberger B, R. Koenker and J. A. Machado, eds., *Economics Applications of Quantile Regression*.
- Chen Lin-an, 1988, "Regression Quantiles and trimmed least squares estimators for structural equations and non-linear regression models", Unpublished PhD. Dissertation, University of Illinois at Urbana-Champaign.
- Chen Lin-an, and Stephen Portnoy, 1996, "Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Structural Equations Models", *Communication in Statistics: Theory and Methods* 25(5): 1005-32.
- Deaton, Angus, 1997, "The Analysis of Household Surveys: A Microeconomic Approach to Development Policy", Baltimore and London: The Johns Hopkins University Press for the World Bank.
- Deolalikar, Anil, 1988, "Nutrition and Labor Productivity in Agriculture: Estimates for Rural South India", *Review of Economics and Statistics*, 70(3): 406-413.
- Fogel R., 1994, "Economic growth, population theory and physiology: the bearing of long-term processes on the making of economic policy". *American Economic Review* 84:369-95.
- García Jaume, Pedro J. Hernández and Angel López-Nicolás, "How wide is the gap? An investigation of gender wage differences using quantile regression", in Fitzenberger B, R. Koenker and J. A. Machado, eds., *Economics Applications of Quantile Regression*.
- Grajeda, Rubén, Jere R. Behrman, Rafael Flores, John A. Maluccio, Reynaldo Martorell, and Aryeh D. Stein, 2005. *The Human Capital Study 2002-04: Tracking, data collection, coverage, and attrition*,

- The Human Capital 2002-04 study in Guatemala: a follow-up to the INCAP longitudinal study 1969-77, *Food and Nutrition Bulletin* 26:2 (Supplement 1), S15-S24.
- Heckman, J., 1974, "Shadow prices, market wages and labor supply", *Econometrica* 42, 679-694.
- Hoddinott, John, Jere Behrman and Reynaldo Martorell, 2005, "Labor force activities and income among young Guatemalan adults," *Food and Nutrition Bulletin* 26:2 (Supplement 1), S98-S109.
- Immink, Maarten D.C., and Fernando E. Viteri, 1981a, "Energy intake and productivity of Guatemalan sugarcane cutters: an empirical test of the efficiency wage hypothesis. Part I", *Journal of Development Economics*, 9: 251-271.
- Immink, Maarten D.C., and Fernando E. Viteri, 1981b, "Energy intake and productivity of Guatemalan sugarcane cutters: an empirical test of the efficiency wage hypothesis. Part II", *Journal of Development Economics*, 9: 273-287.
- Immink, Maarten D.C., Fernando E. Viteri, and Ronald W. Helms, 1982, "Energy intake over the Life Cycle and Human Capital Formation in Guatemalan Sugarcane Cutters", *Economic Development and Cultural Change*, 30(2): 351-372.
- Koenker, Roger and Kevin Hallock, 2001, "Quantile Regression", *Journal of Economic Perspectives*, 15(4): 143-156.
- Koenker, Roger and G. Basset, 1978, "Regression Quantiles", *Econometrica*, 46(1): 1-26.
- Leibenstein, Harvey, 1957, "The theory of underemployment in backward economies", *The Journal of Political Economy* 65(2): 91-103.
- Machado, José A. F. and José Mata, 2001, "Earning functions in Portugal 1982-1994: Evidence from quantile regressions", in Fitzenberger B, R. Koenker and J. A. Machado, eds., *Economics Applications of Quantile Regression*.
- Maluccio, John A., John Hoddinott, Behrman, Jere R., Agnes Quisumbing, Reynaldo Martorell and Aryeh D. Stein, 2005a, "The Impact of an Experimental Nutritional Intervention in Childhood on Education among Guatemalan Adults," Philadelphia-Washington-Atlanta: University of Pennsylvania, IFPRI, Emory, processed.
- Maluccio, John A., Paúl Melgar, Humberto Méndez, Alexis Murphy and Kathryn M. Yount, 2005b, "Social and economic development and change in four Guatemalan villages: demographics, schooling, occupation and assets." *Food and Nutrition Bulletin* 26:2 (Supplement 1), S25-S45.
- Maluccio, John A., Alexis Murphy, Kathryn M. Yount, 2005, "Research note: a socioeconomic index for the INCAP Longitudinal Study 1969-77," *Food and Nutrition Bulletin* 26:2 (Supplement 1).
- Martorell, Reynaldo, 1995, "Results and Implications of the INCAP Follow-up Study", *Journal of Nutrition*, 125 (Suppl): 1127S - 1138S.
- Martorell, Reynaldo, Jere R. Behrman, Rafael Flores, and Aryeh D. Stein, 2005, "Rationale for a Follow-up Focusing on Economic Productivity," *Food and Nutrition Bulletin* 26:2 (Supplement 1), S5-S14.
- Martorell, Reynaldo, Jean-Pierre Habicht, and Juan A. Rivera, 1995, "History and design of the INCAP longitudinal study (1969-77) and its follow up (1988-89)," *Journal of Nutrition*; 125(4S):1027S-1041S.
- Mincer, Jacob B., 1974, "Schooling, Experience, and Earnings", New York, NBER.
- Montenegro, Claudio, 2001, "Wage distribution in Chile: Does gender matter? A quantile regression approach", The World Bank, Development Research Group, Poverty reduction and Economic Management network, Working paper Series N 20.
- Powell, James L., 1983, "The asymptotic Normality of Two Stage Least Absolute Deviations Estimators", *Econometrica*, 51(5): 1569-1576.
- Psacharopoulos, George and Harry Patrinos, 2004, "Returns to Investment in Education: A Further Update," *Education Economics* 12:2, 111-134.
- Ramirez-Zea, Manuel, Benjamin Torun, Reynaldo Martorell and Aryeh D Stein, 2006, "Anthropometric predictors of body fat as measured by hydrostatic weighing in Guatemalan adults," *The American Journal of Clinical Nutrition* 83, 795-802.
- Rosenzweig, Mark R. and T. Paul Schultz, 1985, "The Supply of and Demand for Births: Fertility and Its Life-Cycle Consequence," *American Economic Review* 75(5, December): 992-1015.
- Sahn, David E. and Harold Alderman, 1988, "The Effect of Human Capital on Wages, and the Determinants of Labor Supply in a Developing Country", *Journal of Development Economics* 29(2, September), 157-184.

- Schroeder, Dirk G., Haley J. Kaplowitz and Reynaldo Martorell, 1992, "Patterns and Predictors of Participation and Consumption of Supplement in an Intervention Study in Rural Guatemala," *Food and Nutrition Bulletin* 14(3): 191-200.
- Schultz, T. Paul, 1993, "Investments in the Schooling and Health of Women and Men: Quantities and Returns", *Journal of Human Resources* 28(4): 694-734.
- Schultz, T. Paul, 1997, "Assessing the Productive Benefits of Nutrition and Health: An Integrated Human Capital Approach", *Journal of Econometrics* 77(1): 141-157.
- Stein, Aryeh D., Jere Behrman, Ann DiGirolamo, Ruben Grajeda, Reynaldo Martorell, Agnes Quisumbing, and Usha Ramakrishnan, 2005, "Schooling, educational achievement, and cognitive functioning among young Guatemalan adults", *Food and Nutrition Bulletin* 26:2 (Supplement 1), S46-S54.
- Strauss, John, 1986, "Does Better Nutrition Raise Farm Productivity?", *Journal of Political Economy*, 94: 297-320.
- Strauss, John and Duncan Thomas, 1998, "Health, nutrition, and economic development," *Journal of Economic Literature* 36 (June), 766-817.
- Strauss, John, and Duncan Thomas, 1996, "Wages, Schooling and Background: Investments in Men and Women in Urban Brazil", in Nancy Birdsall, and Richard Sabot, eds., *Opportunity Foregone: Education in Brazil*, 147-91, The Johns Hopkins University Press for the Inter-American Development Bank, Baltimore, MD.
- Strauss, Richard, 2000 "Adult Functional Outcome of Those Born Small for Gestational Age", *Journal of the American Medical Association*, 283(5): 625-632.
- Takeshi, Amemiya, 1982, "Two Stage Least Absolute Deviations Estimators", *Econometrica*, 50(3): 689-712.
- Tannuri-Pianto, Maria E. and Donald M. Pianto, 2002, "Informal employment in Brazil – A choice at the top and segmentation at the bottom: a Quantile Regression Approach", University of Brasilia, Department of Economic Working Paper 236.
- Thomas, Duncan and John Strauss, 1992, "Prices, Infrastructure, Household Characteristics and Child Height", *Journal of Development Economics*, 39(2): 301-331.
- Thomas, Duncan and John Strauss, 1997, "Health and Wages: Evidence on Men and Women in Urban Brazil, *Journal of Econometrics*", 77(1): 159-187.
- Thomas, Duncan, Victor Lavy and John Strauss, 1996, "Public Policy and Anthropometric Outcomes in the Cote d'Ivoire", *Journal of Public Economics*, 61:155-192.
- Thomas, Duncan, John Strauss and M.H. Henriques, 1990, "Child survival, height for age and household characteristics in Brazil." *Journal of Development Economics* 33(2): 197-234.
- Thomas Duncan, and Frankenberg Elizabeth, 2002a, "Health, nutrition and prosperity: a microeconomic perspective". *Bulletin of the World Health Organization* 80(2).
- Thomas, Duncan and Elizabeth Frankenberg, 2002b, "The measurement and interpretation of health in social surveys". In C. J. L. Murray, J. A. Salomon, C. D. Mathers, and A. D. Lopez (Eds.), *Summary Measures of Population Health. Concepts, Ethics, Measurement and Applications*, pp. 387-420. Geneva: World Health Organization.
- Thomas, Duncan, Elizabeth Frankenberg, Jed Friedman, Jean-Pierre Habicht, Mohammed Hakimi, Nicholas Ingwersen, Jaswadi, Nathan Jones, Christopher McKelvey, Gretel Pelto, Bondan Sikoki, Teresa Seeman, James P. Smith, Cecep Sumantri, Wayan Suriastini, Siswanto Wilopo, 2005. "Causal Effect of Health on Labor Market Outcomes: Evidence from a Random Assignment Iron Supplementation Intervention", CCPR On-Line Working Paper 022-04.
- White, Hal, 1980, "A Heteroscedasticity-Consistent Covariance Matrix and a Direct Test for Heteroscedasticity," *Econometrica* 48:4, 817-838.

TABLES

Table 1: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Males						Females				
wage	449	10.907	7.734	1.875	67.103	231	7.933	6.669	0.595	55.833
height	449	162.880	5.816	146.5	180	231	150.579	6.100	118.150	167.150
BMI	449	24.686	3.575	16.197	36.886	231	26.446	4.922	16.816	41.356
educ	449	5.004	3.615	0	16	231	4.502	3.776	0	17
year birth	449	1,970.4	4.000	1,962	1,977	231	1,969.9	4.185	1,962	1,977

Source: Human Capital Study 2002-04 in Guatemala.
See Section 7 for variable definitions.

Table 2: Correlation matrix

	Males					Females				
	wage	height	BMI	educ	year birth	wage	height	BMI	educ	year birth
wage	1					1				
height	0.2266***	1				0.2265***	1			
BMI	0.1833***	0.0232	1			-0.0068	-0.0258	1		
educ	0.4035***	0.2211***	0.1122**	1		0.3388***	0.2445***	-0.1519**	1	
year birth	0.0049	0.0664	-0.1623***	0.1893***	1	-0.0581	0.0176	-0.1154*	0.1578**	1

Source: Human Capital Study 2002-04 in Guatemala.
See Section 7 for variable definitions.

*** statistically significant at 1%; ** significant at 5%; * significant at 10%.

Table 3: Two-step quantile regressions. Dependent variable: ln wages

	Coefficients (std. err.) at quantiles																	
	$\tau = 0.01$		$\tau = 0.05$		$\tau = 0.10$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.90$		$\tau = 0.95$		$\tau = 0.99$	
Guatemalan males, ages 25 to 42 (N = 449)																		
height	-2.2298	(1.9106)	-0.1014	(1.5125)	0.2885	(1.5177)	5.3179	(2.0619)***	4.2666	(1.5586)***	2.9578	(1.6562)*	1.6563	(1.8793)	2.2412	(2.1765)	3.7373 /a	(3.5953)
height²	0.0067	(0.0058)	0.0004	(0.0046)	-0.0007	(0.0046)	-0.0159	(0.0063)**	-0.0128	(0.0047)***	-0.0088	(0.0050)*	-0.0050	(0.0057)	-0.0069	(0.0066)	-0.0115 /a	(0.0109)
BMI	-1.4924	(0.9068)*	-1.0788	(0.5760)*	-0.4126	(0.6588)	0.8273 /a	(0.7708)	1.2197	(0.6506)*	0.8481 /a	(0.7267)	1.7817	(0.8375)**	1.1512 /a	(0.8843)	4.5591	(1.6993)***
BMI²	0.0299	(0.0174)*	0.0199	(0.0112)*	0.0053	(0.0128)	-0.0210 /a	(0.0151)	-0.0274	(0.0129)**	-0.0176 /a	(0.0143)	-0.0330	(0.0164)**	-0.02144 /a	(0.0173)	-0.0923	(0.0332)***
educ	0.0381	(0.0388)	0.0275	(0.0277)	0.0180	(0.0259)	-0.0061	(0.0324)	0.0263	(0.0273)	0.0073	(0.0311)	0.0107	(0.0341)	0.0276	(0.0394)	0.0403	(0.0697)
year birth	-0.0165	(0.0226)	-0.0210	(0.0148)	-0.0272	(0.0131)**	-0.0328	(0.0170)*	-0.0275	(0.0170)	-0.0234	(0.0149)	-0.0035	(0.0209)	0.0130	(0.0245)	-0.0499	(0.0488)
constant	236.3616	(172.9098)	62.3015	(127.0915)	34.3396	(124.5837)	-385.0248	(171.466)**	-311.8382	(129.2634)**	-208.4507	(139.8152)	-151.0550	(166.2491)	-220.0098	(186.8224)	-257.4863	(334.3606)
Guatemalan females, ages 25 to 42 (N = 231)																		
height	-0.1442	(3.0064)	0.4072	(2.3969)	0.5921	(1.9845)	-0.0448	(1.1515)	0.5367	(0.8061)	-0.7637	(1.0754)	-1.3383	(1.5926)	0.4045	(1.9614)	2.7323	(3.0205)
height²	0.0003	(0.0099)	-0.0014	(0.0079)	-0.0020	(0.0066)	0.0002	(0.0038)	-0.0018	(0.0027)	0.0026	(0.0035)	0.0044	(0.0053)	-0.0010	(0.0065)	-0.0086	(0.0100)
BMI	-0.9811	(1.2824)	1.5935	(0.9788)	2.6374	(0.8205)***	0.4033	(0.5259)	0.0002	(0.4244)	0.1641	(0.4658)	0.8185	(0.7450)	1.5439	(0.9297)*	2.8018	(1.2584)**
BMI²	0.0175	(0.0231)	-0.0298	(0.0174)*	-0.0470	(0.0144)***	-0.0085	(0.0094)	-0.0011	(0.0075)	-0.0033	(0.0082)	-0.0151	(0.0129)	-0.0283	(0.0163)*	-0.0522	(0.0224)**
educ	-0.1004	(0.1316)	0.0164	(0.0935)	0.0686	(0.0711)	0.1097	(0.0422)***	0.0746	(0.0326)**	0.0736	(0.0339)**	0.1566	(0.0527)***	0.1168	(0.0633)*	0.2261	(0.0951)**
year birth	-0.1028	(0.0474)**	-0.0293	(0.0427)	-0.0291	(0.0364)	-0.0280	(0.0270)	-0.0228	(0.0182)	-0.0224	(0.0184)	-0.0337	(0.0292)	-0.0457	(0.0413)	-0.1442	(0.0634)**
constant	231.8344	(213.7229)	8.2030	(187.4151)	-22.4114	(155.4234)	54.9492	(87.3396)	7.5158	(64.1639)	100.9004	(81.6984)	158.1562	(120.0329)	34.2508	(141.9069)	33.6612	(207.3621)

Notes: Height, height sqrd, BMI, BMI sqrd and educ instrumented with atole largevill largevilla byr twins water_02 presch_02 exp00_24 exp00_24a exp24_42 exp24_42a pstruc_07 stu_tch_07 dthma_12 dthpa_12 SES momeduc dadeduc dummomeduc dumdadeduc. See Section 7 for variable definitions.

Bootstrapped standard errors (5000 replications) in parenthesis

*** significant at 1%; ** significant at 5%; * significant at 10%.

/a These coefficients are statistically significant at 5% under the bias-corrected confidence interval method.

Table 4: Test of equality of returns to health for quantile regression estimates with instrumental variables

Quantiles		Height	Height ²	BMI	BMI ²
		<i>p-value</i>			
0.01	0.05	0.245	0.260	0.618	0.532
0.01	0.10	0.234	0.248	0.231	0.157
0.01	0.25	0.005	0.005	0.034	0.015
0.01	0.50	0.008	0.009	0.013	0.006
0.01	0.75	0.047	0.051	0.046	0.036
0.01	0.90	0.168	0.173	0.007	0.008
0.01	0.95	0.137	0.139	0.042	0.041
0.01	0.99	0.154	0.154	0.003	0.002
0.05	0.10	0.739	0.748	0.143	0.102
0.05	0.25	0.006	0.007	0.010	0.005
0.05	0.50	0.028	0.028	0.003	0.002
0.05	0.75	0.181	0.185	0.032	0.033
0.05	0.90	0.499	0.495	0.006	0.009
0.05	0.95	0.414	0.404	0.045	0.057
0.05	0.99	0.340	0.332	0.002	0.002
0.10	0.25	0.001	0.001	0.039	0.025
0.10	0.50	0.022	0.023	0.022	0.020
0.10	0.75	0.228	0.230	0.155	0.182
0.10	0.90	0.605	0.598	0.035	0.059
0.10	0.95	0.502	0.489	0.167	0.224
0.10	0.99	0.394	0.383	0.007	0.007
0.25	0.50	0.501	0.517	0.504	0.582
0.25	0.75	0.314	0.323	0.981	0.841
0.25	0.90	0.206	0.216	0.388	0.582
0.25	0.95	0.317	0.336	0.785	0.986
0.25	0.99	0.702	0.725	0.045	0.050
0.50	0.75	0.393	0.394	0.558	0.436
0.50	0.90	0.237	0.245	0.552	0.763
0.50	0.95	0.412	0.431	0.947	0.770
0.50	0.99	0.888	0.907	0.056	0.058
0.75	0.90	0.424	0.438	0.231	0.313
0.75	0.95	0.744	0.770	0.744	0.834
0.75	0.99	0.830	0.812	0.028	0.024
0.90	0.95	0.723	0.705	0.364	0.393
0.90	0.99	0.550	0.540	0.094	0.067
0.95	0.99	0.642	0.639	0.032	0.022

Note : This table corresponds to tests of equality of returns to health across quantiles for the instrumented model in Table 3 for males. P-values are based on a χ^2 test.

FIGURES

Figure 1

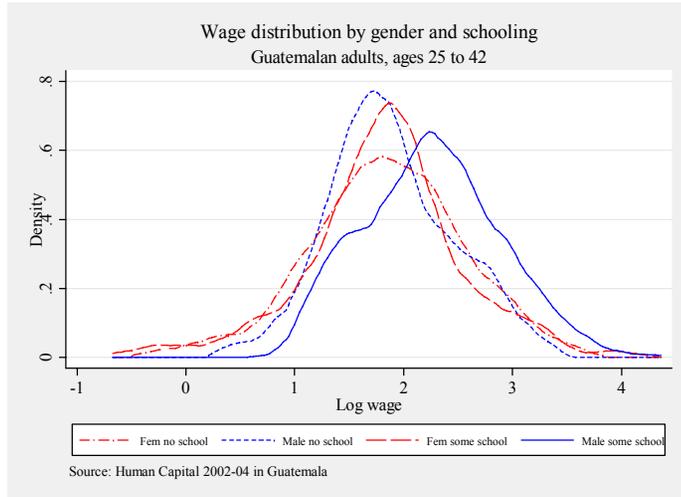


Figure 2

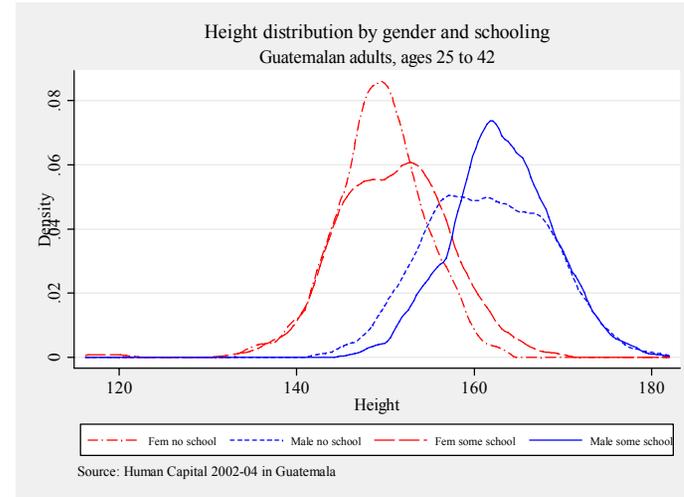


Figure 3

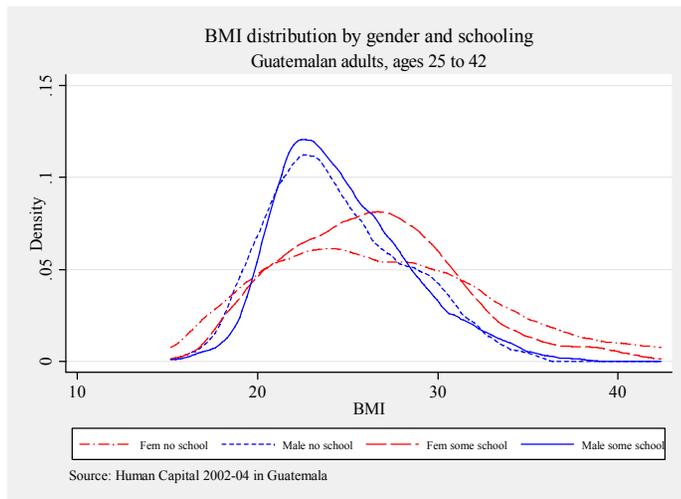


Figure 4

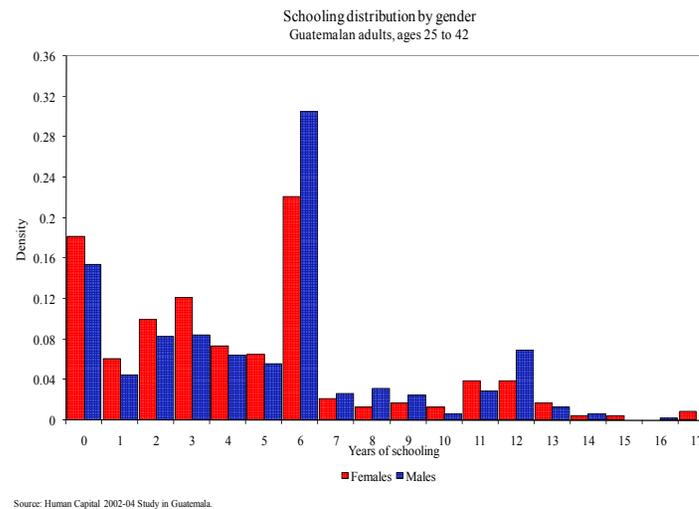


Figure 5

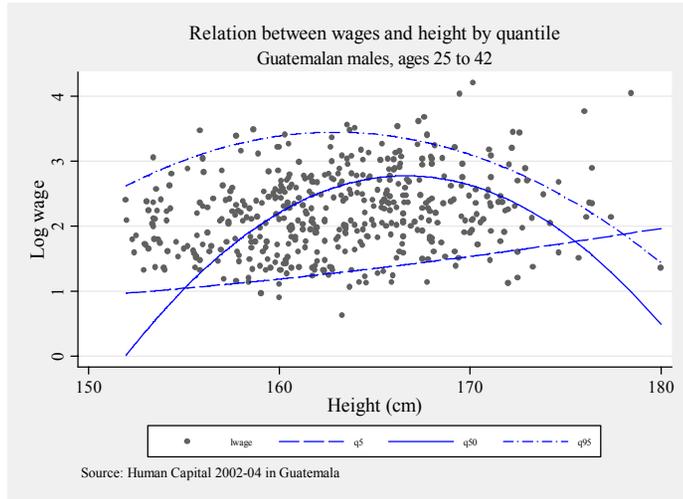


Figure 6

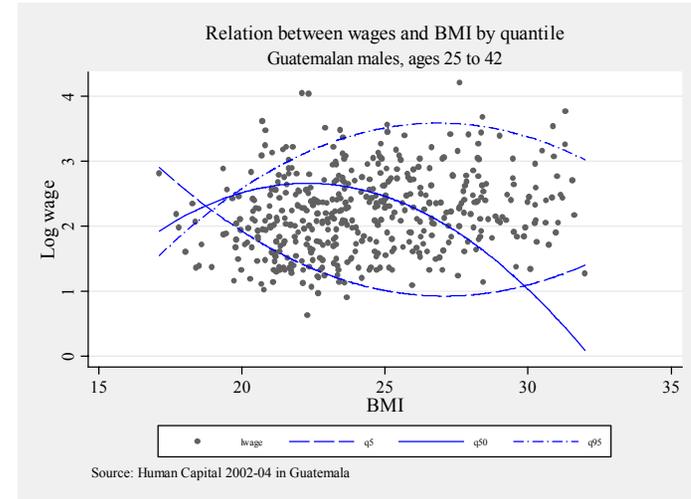


Figure 7

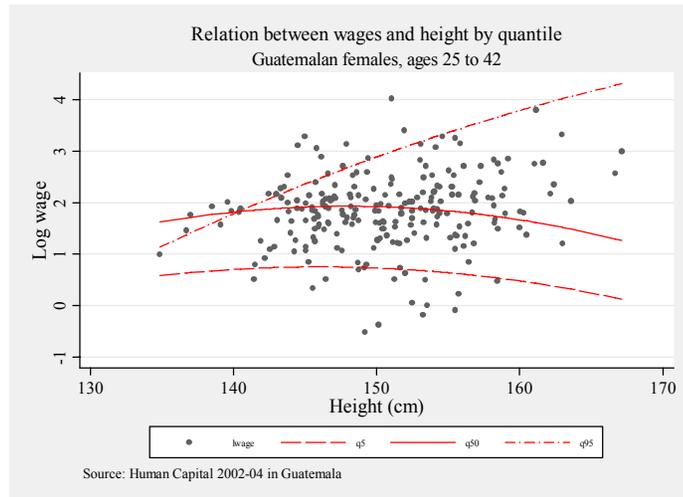


Figure 8

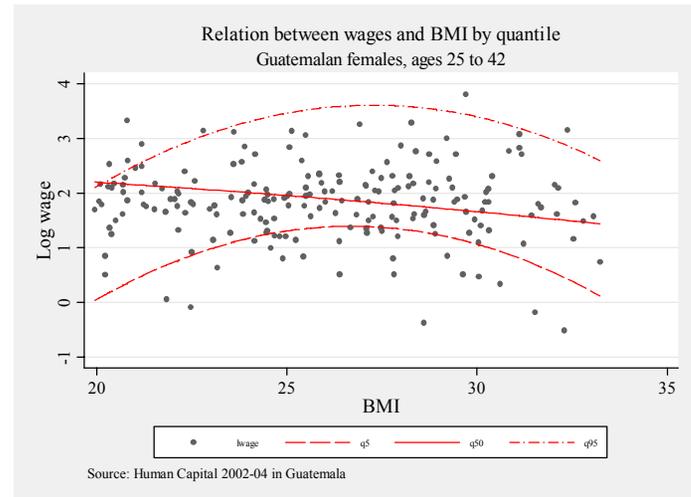


Figure 9

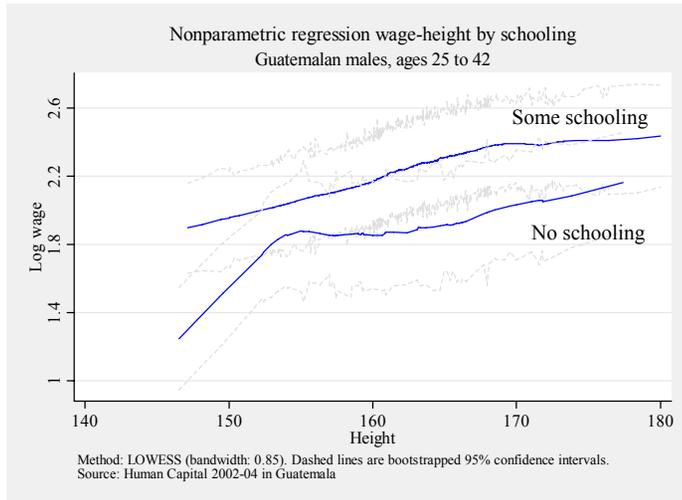


Figure 10

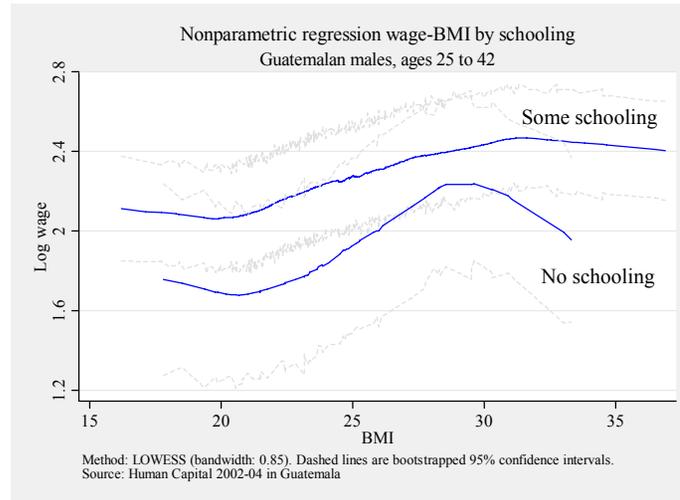


Figure 11

Relation between Productivity, BMI and Education (males)

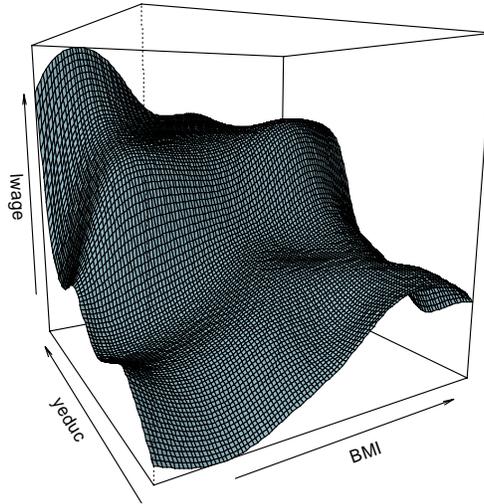


Figure 12

Relation between Productivity, Height and Education (males)

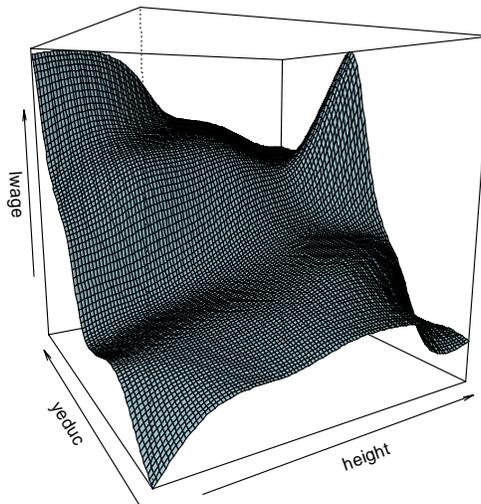


Figure 13

Relation between Productivity, Height and BMI (males)

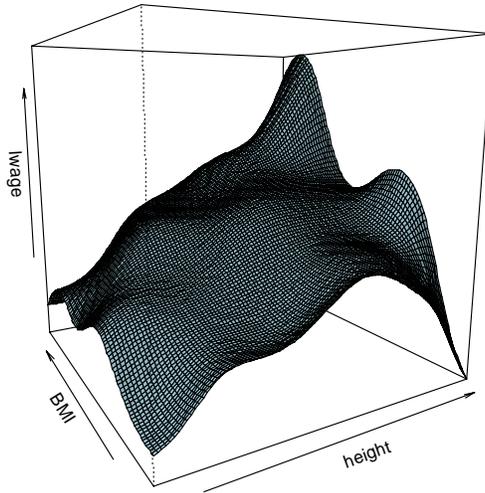


Figure 14

Relation between Productivity, BMI and Education (females)

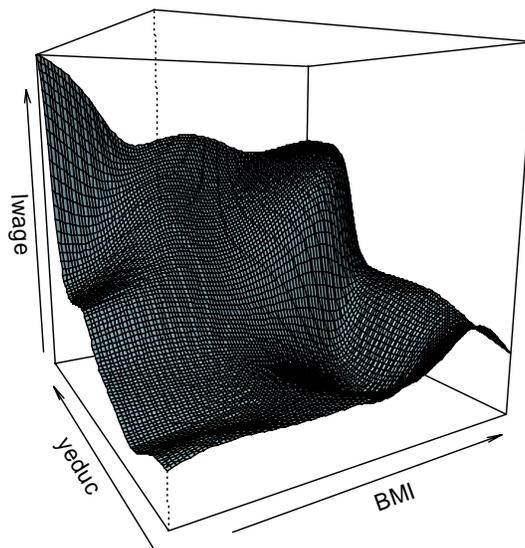


Figure 15

Relation between Productivity, Height and Education (females)

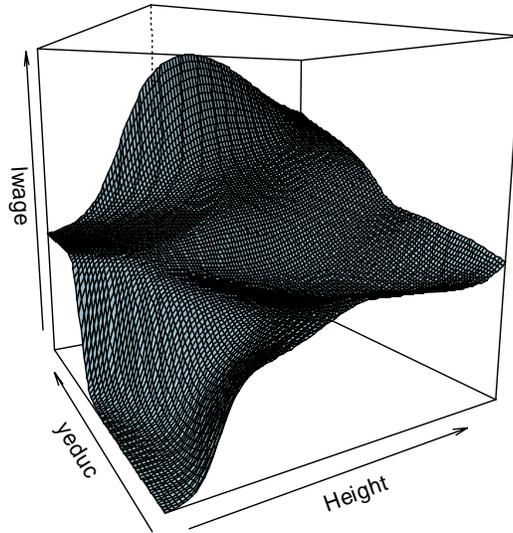


Figure 16

Relation between Productivity, Height and BMI (females)

