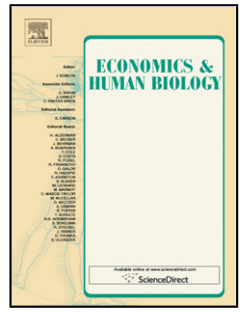
Accepted Manuscript



Title: Combining insights from quantile and ordinal regression: Child malnutrition in Guatemala

Authors: Stuart Sweeney<!–<query id="Q1">Please confirm that given names and surnames have been identified correctly.</query>–>, Frank Davenport, Kathryn Grace

PII: S1570-677X(12)00076-7
DOI: doi:10.1016/j.ehb.2012.06.001
Reference: EHB 396

To appear in: *Economics and Human Biology*

Received date: 10-6-2011
Revised date: 26-5-2012
Accepted date: 1-6-2012

Please cite this article as: Sweeney, S., Davenport, F., Grace, K., Combining insights from quantile and ordinal regression: Child malnutrition in Guatemala, *Economics and Human Biology* (2010), doi:10.1016/j.ehb.2012.06.001.

Combining insights from quantile and ordinal regression: Child malnutrition in Guatemala

Stuart Sweeney[a,*], Frank Davenport[a], Kathryn Grace[b]

[a]Department of Geography, University of California, Santa Barbara CA 93106-4060
[b]Department of Geography, University of Utah, Salt Lake City UT 84112-9155

Abstract

Chronic child undernutrition is a persistent problem in developing countries and has been the focus of hundreds of studies where the primary intent is to improve targeting of public health and economic development policies. In national level cross-sectional studies undernutrition is measured as child stunting and the goal is to assess differences in prevalence among population subgroups. Several types of regression modeling frameworks have been used to study childhood stunting but the literature provides little guidance in terms of statistical properties and the ease with which the results can be communicated to the policy community. We compare the results from quantile regression and ordinal regression models. The two frameworks can be linked analytically and together yield complementary insights. We find that reflecting on interpretations from both models leads to a more thorough analysis and forces the analyst to consider the policy utility of the findings. Guatemala is used as the country focus for the study.

Keywords: child chronic undernutrition, quantile regression, ordinal regression, Guatemala

1. Introduction

Malnutrition and starvation are enduring interests in the development literature, especially in the ‘basic needs’ conception of development. Malnutrition concerns development policy because nutritional status is often strongly associated with productivity, labor force participation, and educational attainment, especially in poor agrarian societies (Berg and Muscat, 1972; Alderman et al., 2006; Strauss and Thomas, 1998; Jamison, 1986).

*Corresponding author. Phone: 805-618-8317
Email addresses: sweeney@geog.ucsb.edu (Stuart Sweeney), frank.davenport@gmail.com (Frank Davenport), grace@geog.ucsb.edu (Kathryn Grace)

Preprint submitted to Elsevier *June 10, 2012*

In this paper we evaluate the science and policy interpretations of regression models used to study child malnutrition generally, and specifically as applied to the context in Guatemala. We focus on chronic malnutrition and use a widely accepted *stunting* metric as an operational measure. Much of the research and policy literature has relied on standard linear regression or binary outcome models. Several recent papers have used ordinal regression models and a handful of academic studies have utilized quantile regression. As compared to standard regression, these conceptually similar models relax constraints in the model specification permitting more targeted interpretations of covariate effects. We draw analytic connections between the two specifications, with the goal of examining the usefulness of an integrated interpretation of chronic child malnutrition.

The integration of the model results is primarily motivated by the potential advantages in clarifying the policy interpretations. The comparison also highlights complications that arise in the ordinal regression model estimation. The utility of quantile regression will only be of interest under location-scale shifts, but that implies a heteroskedastic error distribution and inconsistent ordinal regression estimates. We recover consistent estimates with a two-stage estimator that directly links information about the error distribution from a first stage median regression to the second stage ordinal regression.

The model comparison is framed using recent data from Guatemala. Plagued by the highest rates of chronic undernutrition in Latin America – the prevalence of stunting among Guatemalan children under 5 years of age is nearly 55% – health and policy researchers continue to explore the most effective means for improving the nutrition of Guatemalans (Espindola et al., 2005; WHO, 2010). As recent malnutrition research indicates stagnant, and possibly worsening, levels of chronic undernutrition (WHO, 2010) identifying the policy-relevant factors and their potential impact on stunting continues to be an important area of research. In this study we develop model specifications motivated by past research of child stunting in Guatemala and broadly informed by general food security and nutrition health and policy literatures. Our intent is not to produce a definitive analysis of Guatemalan child undernutrition. Rather, we use the Guatemalan context to demonstrate how the integrated modeling strategy can yield useful policy insights.

The paper proceeds as follows. Section 2 explores the policy context and conceptualizations underlying malnutrition research and then reviews the standard model specifications that have been used in past empirical studies. Quantile regression and ordinal regression models are reviewed in detail. Section 3 describes the data source and defines a consensus set of independent variables from published research on Guatemala. Sections 5, 6 and 7 contain the results, a discussion, and then conclusions.

2. Malnutrition: Concepts, Models, and Policy Interpretations

Malnutrition refers to any deviations from normal "healthy" nutrition; both undernutrition and over-nutrition. In this study we focus on under-nutrition which is the more common type of malnutrition experienced by children in Guatemala. While detailed clinical tests and diet analysis can isolate the causes and nature of undernutrition at the individual level, anthropometric measures were developed for large scale prevalence studies that can be implemented in countries even when the national public health infrastructure is weak or non-existent (Seoane and Latham, 1971). The standard anthropometric measure of chronic child undernutrition is the height (length)[1] for age Z-score (HAZ)(WHO, 2004).[2] The Z-score is derived by comparing an individual or population to the international standard height growth curves that were developed in the Multicenter Growth Reference Study (MGRS)[3], conducted by the World Health Organization from 1997 and 2003 (WHO, 2004).[4] Children are considered healthy if their HAZ score lies within two standard deviations of the mean height for age ratio ($|HAZ| \leq 2$). Undernutrition is suspected when a child's score is more than two standard deviations below the mean. If $-3 < HAZ \leq -2$ children are classified as *stunted* (or malnourished) and if $HAZ \leq -3$ children are considered *severely stunted* (or severely malnourished). Guatemala's extreme levels of chronic undernutrition are reflected in Figure 1. In Guatemala, the distribution of HAZ for children aged 12-36 months is shifted to the left so that approximately 49% of the population is either *severely stunted* or *stunted*.

2.1. Models

Past research on stunting has used standard linear regression (Pebley and Goldman, 1995; Marini and Gragnolati, 2003), ordered or unordered categorical response models (Sahn and Stifel, 2002; Brennan et al., 2004; Lee et al., 2010), or quantile regression (Borooah, 2005; Sturm and Datar, 2005; Bassolé, 2007; Aturupane et al., 2008; Fenske et al., 2009; Kandpal and McNamara, 2009). These methods are typically presented in isolation and as the *best* or most appropriate specification for building empirical models of malnutrition. We explore an alternative approach in this paper where we combine the insights gained from both quantile regression and ordinal regression.

[1]Length is used for children younger than 24 months, height is used for children older than 24 months (WHO, 2004).

[2]Height for age is the preferred measure of chronic malnutrition because it is less likely than other indicators to be impacted by disease or other sources of stress when the data is collected. For example Borooah (2005), citing Sahn and Stifel (2002), notes that height for age, unlike weight, will not be impacted by temporary ailments such as diarrhea or malaria.

[3]http://www.who.int/childgrowth/mgrs/en/

[4]The length for height Z-scores used in this study were calculated using the WHO Anthro 2.02 macros for SPSS, see http://www.who.int/childgrowth/software/en/.

Quantile regression is easiest to understand in relation to linear regression estimated using ordinary least squares (OLS). In OLS, we estimate the parameters, β, by minimizing the sum of squared deviations from the conditional mean of the dependent variable ($\sum(y_i - \hat{y}_i)^2$, where $\hat{y}_i$ is the conditional mean $\mu(x_i, \beta)$). The resulting parameters are interpreted as the shift in the conditional mean of y given a unit change in x. Quantile regression simply changes the distance function and the pivot point used in the minimization problem (Koenker, 2005; Koenker and Bassett, 1978).[5] The parameter estimates are interpreted as a shift in the location of quantile, τ, given a unit change in covariate x. As such, the parameter estimates may vary depending on the quantile chosen:

$$Q_y[\tau|x_i] = \beta_0(\tau) + \beta_1(\tau)x_i.$$

Based on the discussion of our dependent variable (HAZ), we chose $\tau = \{0.26, 0.49, 0.91\}$. The first two correspond to the approximate thresholds for severe stunting and stunting, and the last is approximately centered in the "healthy" range for Guatemala and is near $y_i = 0$.

For ordinal regression models we convert the continuous variable, y_i, into an ordered categorical variable, y_i^*, with the levels ($j = 1, 2, 3$) using the quantile thresholds defined above:

$$y_i^* = \begin{cases} 1 & \text{if } y_i \leq -3 \\ 2 & \text{if } -3 < y_i \leq -2 \\ 3 & \text{if } y_i > -2 \end{cases}$$

Ordinal regression is used to describe variation among individuals in the probability of class membership ($j = 1, 2, 3$) conditional on covariates, $Pr(y_i^* \leq j|x_i) = F(x_i^T \gamma_j)$.[6] The direct parameter estimates are difficult to interpret for ordinal models but post-processing allows us to recover the marginal effects of covariates on unconditional probabilities $Pr(y_i^* = j)$ and conditional probabilities $Pr(y_i^* = k|k \geq j)$ (Harrell, 2001; Gelman and Pardoe, 2007). The marginal effects for the latter probabilities are somewhat easier to interpret. For the sake of completeness, we present both results in section 4.

Quantile and ordinal regression are approximately linked through an inverse relationship. Assume $y^{(t)}$ is a threshold value of the random variable $\mathbf{Y}$ with distribution function, $F : \mathbb{R}_y \rightarrow \tau \in [0, 1]$. Any threshold $y^{(t)}$ has an associated and unique probability $\tau^{(t)}$ if F is a monotonic increasing function; $F(\mathbf{Y} \leq y^{(t)}) = \tau^{(t)}$.

[5]For a given quantile, β is estimated by minimizing, $\sum_{i=1}^{n} \rho_\tau(y_i - \xi(x_i, \beta))$, where the check function with argument v is defined,

$$\rho_\tau(v) = \begin{cases} \tau v & \text{if } v \geq 0 \\ (1 - \tau)v & \text{if } v < 0 \end{cases}$$

[6]The coefficients are given the non-standard symbol γ rather than β to distinguish between the two sets of estimates.

In quantile regression, the inverse distribution provides the mapping, $G(F(\mathbf{Y} \leq y^{(t)})) = G(\tau^{(t)}) = y^{(t)}$. Given a fixed set of quantiles $\tau^{(t)}$ we study variation in $y^{(t)}$ conditional on covariates. Ordinal regression operates directly in the distribution function space for ordered discrete outcomes, y^*, with probabilities $Pr(y^* = j) = \int_{y_k^{(t)}}^{y_{k+1}^{(t)}} f(u)du$ where $y_k^{(t)}$ and $y_{k+1}^{(t)}$ are two threshold values. In ordinal regression the threshold values $y^{(t)}$ are held fixed and we study variation in the probabilities $Pr(y^* = j)$ conditional on covariates. As such, the outcomes for the two models are mathematically related but without knowing the exact functional form of F it is only a loose coupling, and in quantile regression the functional form of F is left unspecified.[7] As long as the underlying empirical CDF is increasing and the ordinal regression parameters are estimated consistently, then interpretations from the two models provide complementary perspectives on the same underlying process.

2.2. Estimation and specification testing

For the quantile regression parameter estimation we use well-established methods that are available in major software packages (Koenker, 2010).[8] In addition to general specification testing there is particular interest in assessing whether there is evidence of changes beyond a simple conditional shift in the location of the quantile. There are several descriptive approaches and formal tests that can be used to characterize changes in the response distribution. These include testing constant slope hypotheses ($H_0 : \beta(0.26) = \beta(0.49) = \beta(0.91)$), interpreting ordering among quantile effects, and direct tests against a null of *pure location shift* or a null of *location-scale shift* (Hao and Naiman, 2007; Koenker and Xiao, 2002; Handcock and Morris, 1999). If all, or most, of the estimated effects provide no evidence refuting a pure location shift then the additional flexibility afforded by quantile regression is unnecessary, and conditional mean models should suffice. In most of the published papers using quantile regression to study child stunting or other anthropometry measures there is evidence of location-scale shifts or more general evidence of complex conditional distributions. Given that microdata from survey samples is typically characterized by heteroskedasticity (Greene, 2003), the existence of location-scale shifts comes as no surprise. In these cases quantile regression estimates remain consistent but there is a slight loss in efficiency.

In the presence of heteroskedasticity, ordinal regression models will yield inconsistent coefficient estimates (Greene, 2003). Consider a simple binary model for the probability of severe stunting. The underlying

[7]Quantile regression uses only the order statistics of $\mathbf{Y}$.

[8]We coded our own pairs cluster bootstrap and the standard errors and test statistics are based on the bootstrap covariance matrix. Results based on simple pairs bootstrap, a weighted pairs bootstrap, and a weighted cluster bootstrap are available from the authors.

index function, $y_i = x_i^T\gamma + \epsilon_i$, with $y^* = 1$ if $y_i < -3$ and $y^* = 0$ otherwise is:

$$Pr(y_i < -3|x_i) = Pr(x_i^T\gamma + 3 + \epsilon < 0) = Pr(\epsilon > x^T\gamma^*).$$

The intercept absorbs the +3 shift (indicated by the $*$ on γ^*) and captures the non-central probability evaluation. Under heteroskedasticity of an unknown functional form, $\sigma(x)$, the index function becomes, $y_i = x_i^T\gamma + \sigma(x_i)\epsilon_i$, the probability model is,

$$Pr(y_i < -3|x_i) = Pr\left(\epsilon_i > \frac{x_i^T\gamma^*}{\sigma(x_i)}\right),$$

and MLEs of γ^* based on $\mathbf{x}$ are inconsistent (Greene, 2003).

Typically when estimating ordinal regression models the latent continuous dependent variable y_i is unobserved. In our case we observe y_i and we can use residuals from a first stage estimation to develop an estimate of the scaling factor $\sigma(x_i)$. Specifically, we adapt Zhao's (2001) approach to consistent median regression estimation under heteroskedasticity of an unknown form. His approach builds on work by Newey and Powell (1990) that relies on Stone's (1977) k nearest neighbors (k-NN) estimators applied to the residuals from a median regression of the untransformed data. The basic estimator has the form:

$$\hat{\sigma}_i = \sum_{j=1}^{n} W_{ij}|\hat{e}_{ij}| \quad i = 1, 2, \ldots, n \tag{1}$$

where W_{ij} are k-NN weights. Zhao's (2001) paper contains proofs of consistency and Monte Carlo simulation results for the median regression case that demonstrate the relative efficiency of alternative estimators. Also note that the estimated scaling effects ($\hat{\sigma}_i$) no longer depend on observed or unobserved covariates. In our application, we use a parallel estimation strategy, except that in the second stage we transform the design matrix by dividing each row by, $\hat{\sigma}_i$, prior to estimating the ordinal regression, and our concern is improving relative consistency.[9]

The full set of thresholds and covariate effects, $\hat{\gamma}_j$, in the ordinal regression are estimated jointly in a single model specification. The binary outcome variable is the $2n \times 1$ vector $\mathbf{y}^* = [I(\mathbf{y} \leq -3), I(\mathbf{y} \leq -2)]$,

[9]A simulation study of the two-stage estimators properties is available from the authors. The Monte Carlo simulation parallels section 3 of (Zhao, 2001). We simulate data with three types of heteroskedasticity. In stage 1, we fit a median regression and used the residuals from the model to estimate $\hat{\sigma}_i$. The k-NN weights based on 30% of the sample and a triangular kernel were then used to transform regressors and intercepts in the ordinal model. The distribution of the estimates were compared to a model with no heteroskedasticiy, the uncorrected index function, and an index function transformed by the known weights. The untransformed model yielded clearly biased parameter estimates. The k-NN based weights removed substantial bias and were very close to the true model and almost identical to model based on an index function transformed by the true weights. While the the rescaling by $\hat{\sigma}_i$ removes bias, there is an accompanying marked degradation in efficiency.

where I is an index function taking the value 1 if the condition is true and 0 otherwise. The $2n \times 2(k+1)$ design matrix for the regression is,

$$\mathbf{X}^* = \mathbf{I}_2 \otimes (diag(\hat{\sigma}^{-1})\mathbf{X}) \tag{2}$$

where $\mathbf{I}_2$ is a 2×2 identity matrix, σ^{-1} is a vector with elements $\{1/\hat{\sigma}_i\}$, and $\mathbf{X}$ is an $n \times (k+1)$ design matrix (including intercept) for k covariates.[10] Standard errors are calculated using a pairs cluster bootstrap.[11]

A final issue with the ordinal regression model is interpretation of the parameter estimates. Direct interpretation of ordinal regression parameter estimates is more complicated than interpretation for simple binary outcome logit or probit models (Greene, 2003). We follow the recommendation of Gelman and Pardoe (2007) and report average predictive comparisons on the probability scale. Essentially, the idea is to use the inverse logit transformation to calculate pairwise differences in predicted probabilities over the full data set while also incorporating uncertainty in parameter estimates. The resulting average predictive comparisons for each effect have an interpretation similar to standard regression analysis; it is the change in the probability expected from a one unit change in the covariate effect with other covariates held fixed. In our application it also has the domain specific interpretation of the average model predicted difference in prevalence from a one unit change in the covariate. Another advantage of using average predictive comparisons is that probability predictions can be combined to form more interpretable marginal probability effects. In our case the direct model predictions yield probabilities of severe stunting, $Pr(y^* = 1|x)$ and of severe stunting or stunting, $Pr(y^* \leq 2|x)$. Within the average predictive comparison framework we can recover the unconditional probabilities $Pr(y^* = 2|x)$ from $Pr(y^* \leq 2|x) - Pr(y^* = 1|x)$, $Pr(y^* = 3|x)$ from $1 - Pr(y^* \leq 2|x)$, and the conditional probability, $Pr(y^* = 2|y^* > 1, x)$ using $[Pr(y^* \leq 2|x) - Pr(y^* = 1|x)]/[1 - Pr(y^* = 1|x)]$.

To sum up, the estimation strategy we propose is as follows:

1. Fit a quantile regression and use formal and informal tests to assess whether there are complex distributional features. If not, analysis can proceed with OLS and/or standard ordinal regression.

[10] k is inclusive of the dummy variable created for Region and other multi-category qualitative covariates. The design matrix approach used to simultaneously estimate multiple thresholds in a cumulative logit is discussed originally in Winshop and Mare (1984).

[11] An R code implementation of the two stage estimation: 1) k-NN weighting estimation of σ_i from $\hat{e}_i$, and 2) ordinal regression with various bootstrap estimator options, is available from the lead author.

2. Fit a median regression, recover the residuals, and use (1) to estimate the scaling effects, $\hat{\sigma}_i$.
3. Construct $\mathbf{y}^*$, $\mathbf{D}^*$ from (2), and then fit an ordinal logit or probit model of $\mathbf{y}^*$ on $\mathbf{D}^*$.
4. Use average predictive comparisons to recover probabilities that are most relevant for the problem under study.

2.3. Policy interpretations

Studies of undernutrition using regression based approaches are intended to support and guide nutritional planning and to promote the importance of nutrition in broader development policy. If the studies use anthropometric measures the goal is typically to characterize and compare prevalences among subpopulations. How exactly should the results of these models be interpreted in this policy context and how are those interpretations supported by integrating quantile and ordinal regression results?

While the underlying anthropometric measures of undernutrition are continuous, in the policy context the concept of chronic undernutrition is most frequently expressed as categorical. As noted previously, the World Health Organization defines *stunting* and *severe stunting* as categories in relation to standard height-for-age z-scores. Nutrition policy therefore tends to mirror this categorical classification; assessment of national or regional chronic undernutrition is typically presented as the proportion of the population that is classified as stunted or severely stunted. Similarly, nutrition program development goals are conceived of in terms of reducing the share of population in those categories.

Ordinal regression models provide a direct conceptual link to the discrete category framing of chronic undernutrition. Using transformations to the probability scale and average predictive comparisons, ordinal models allow covariate effects to be interpreted as the probable change, given a unit change in the covariate, in the proportion of the population that is either *severely stunted*, *stunted*, or *stunted conditional on not severely stunted*. In a policy setting, the results can be linked to the categorical framing and used to develop targeted outreach. The predictive differences from the ordinal regression models can be viewed as smoothed versions of the underlying data that highlight specific contrasts.

While the policy merits of direct estimates of changes in prevalence are relatively self-evident, the prevalence estimates do not provide a complete picture for policy interpretation. In particular, there is little information provided about the potential efficacy of targeting a particular group, the exact nature of the targeting that should be pursued, or the potential trade-offs that should be considered.

The information about conditional shifts in distributional form from quantile regression can actually speak to some of those issues. Consider a simple binary variable effect that results in a leftward shift in the distribution (towards higher prevalence of stunting or severe stunting). If the scale change is also negative, it implies the subpopulation is relatively more homogeneous in being worse off, and in turn implies that policies targeting the subpopulation will be more effective compared to a case when the scale does not decrease. The exact changes in prevalence will not be clear from the quantile results and will depend on the size of the location shift, but this is exactly where the ordinal regression results can complement the interpretation.

A second case to consider is when the leftward shift in the distribution is also characterized by skewing to the left. In this case, the skewing might reveal that there is a small population of extreme outcomes within the severe stunting category, for example. While the prevalence change might reveal an increase, the additional information about the presence of extreme cases should inform the policy process. For example, policies could be pursued to target the most extreme cases that would be judged successful even if the prevalence of extreme child stunting does not improve. It would also be justification for fitting ordinal models with more tail categories. But those models might have too few data points to effectively capture the pattern.

Those are just the two most obvious cases when distributional effects made evident by quantile regression would provide useful information to policy assessment and development. Given the broad range of potential distributional effects that might be identified we are suggesting that each of those may prove useful in a policy setting, and that those insights would be missed using conditional mean models. We believe that the joint consideration of insights about covariate effects on conditional distributions with covariate effects on changes in prevalence will force analysts to engage with the real complexities of the processes under study, and that communication of the results will resonate with a policy audience because they are framed using the benchmarks and measures dominant in national and international health planning.

We should note that in most published studies, the quantiles chosen are a systematic selection of the domain of y – such as all deciles or all quintiles (Abrevaya, 2001; Borooah, 2005; Bassolé, 2007). To identify the kinds of distributional effects introduced above requires at least three quantiles. While more nuanced information could be gained from producing results at a greater number of quantiles, perhaps every 5th or 10th percentile, our goal in presenting the analysis of the Guatemalan data is to provide an example of the interpretations discussed in this section using real survey data. The choice of three quantiles allows for more simple tabular and graphical comparisons while still allowing us to capture changes in conditional distributions. Moreover, we also believe that analysis of quantiles that are near the thresholds of $HAZ = \{-3, -2\}$ (mapping to $\tau = \{0.24, 0.49\}$ in our study) provides a meaningful connection to the ordinal regression analysis. If we had chosen to include additional quantiles, we would have retained $\tau = \{0.24, 0.49\}$ in the broader set.

3. Data

To demonstrate the proposed advantages in policy interpretation we provide an example analysis and interpretation focused on child undernutrition in Guatemala. The data are from the 2000 Guatemala Living Standards and Measurement Survey (LSMS). The LSMS is an in-depth household, community, and price survey; the household portion covers 7,276 households with a total of 37,771 individuals.[12] The survey is nationally representative and includes modules for demographic, economic, health, and community data (INE, 2000). We selected the subset of records for children aged 12 to 36 months. If a household had multiple children in the age range, then we selected only the youngest child to avoid estimation issues associated with household-level clustering. Our final sample used in this analysis contains information for 1,520 Guatemalan children (see Table 1).

As mentioned in the introduction, the purpose of this paper is not to revisit previous work on child malnutrition in Guatemala, but rather to highlight the advantages of the modeling strategy outlined in section 2. Thus our choice of variables is not exhaustive but is reflective of the types of measures used in published empirical studies of Latin American malnutrition (Marini and Gragnolati, 2003; Pebley and Goldman, 1995; Farrow et al., 2005; Larrea and Kawachi, 2005; Balk et al., 2005).

We rely on parents' height (average of the mother's and father's height), child's age and the sex of the child to capture biological-based variation. The household and parental variables used are ethnicity, [13] a breastfeeding measure, mother's education, a poverty indicator, floor material, and the number of children in the household under age five. We also include regional and community effects that have been highlighted

[12] The data was collected by the Guatemalan National Statistics Institute (INE), with assistance and technical guidance from the World Bank LSMS team, from July to December 2000 . A total of 745 primary sampling units (PSU's) were selected at random from 11,170 spatial sectors (similar to USGS 7 quadrangles). The sectors were pre-designated as rural or urban based on results from a 1994 population census. A roughly equal number of sectors was selected for each of the eight administrative regions in Guatemala.

[13] Indigenous people differ from their non-Indigenous counterparts in language (there are 22 non-Spanish languages spoken in Guatemala) and often in dress. Indigenous people are generally economically, socially, and politically marginalized and face many barriers (language, economic, educational) to health and public services (Pebley and Goldman, 1995; Lee et al., 2010)

as significant in related studies. Specifically we include type of place of residence (urban versus rural), region of residence, and community water source.

4. Results

The results in this section are provided as an example of the range of distributional effects that can be detected in analysis of survey data, and as a demonstration of the interplay between interpretations of quantile and ordinal regression models. The model results are provided in two forms. Tabular displays of the coefficients and standard errors can be found in Tables 2 and 3. These same results are presented graphically in Figures 3, 4, and 5. Both models produce a fairly large and complex set of coefficients, and the figures were designed to facilitate comparison of the relative magnitudes and dispersion across τ's for a single variable, across variables, and between the models. The dispersion is represented by horizontal lines spanning the 10th to 90th quantiles of the parameter distribution. The major difference between the graphical results is that the quantile effects are on the scale of HAZ whereas the ordinal regression results are the expected change in probability, or equivalently the predicted change in prevalence. As such, beneficial outcomes are characterized by positive coefficients in the quantile results (representing a rightward shift in the conditional quantile) and negative coefficients in the ordinal results (the probability of being classified as stunted decreases).

Following the analytic strategy proposed above, we start by assessing whether the quantile results provide evidence of distributional complexities beyond a pure conditional location-shift. An informal assessment is to check whether parameter estimates are uniformly increasing or decreasing with quantiles and whether those differences are significant. Comparing the pairwise tests of fixed slopes in Table 4 and effect ordering in Figure 3, the effects for urban, Southwest region, Northwest region, Indigenous, and bare floor are all suggestive of location-scale shifts or more complex distributional effects. Notice that several other effects, for example parents' height or number of children, would be captured adequately with a pure location-shift/conditional mean effect. A more formal approach is to use hypothesis tests against a null of pure location shift or a null of location-scale shift (Koenker and Xiao, 2002). One effect (Southwest) rejected the null of pure location shift at the α =0.10 level and three others (Northeast, age, and water delivery) rejected it at the α =0.15 level. Four effects (Northeast, Southwest, water delivery, and education) rejected the null of location-scale shift at the α =0.10 level.[14] Overall, we have sufficient evidence to

[14]Critical values are 2.57 (α =0.01), 2.05 (α =0.05), and 1.77 (α =0.10). The test statistics for location-shift are Southwest

assume that the more robust and flexible quantile specification is warranted. As noted in the methodology section, this in turn has implications for the ordinal regression, and specifically argues for the use of our heteroskedasticity-adjusted estimator.

As a specific example of integration of the two regression approaches and their advantages over simple OLS or binary models, we explore in detail the effects for the binary ethnicity covariate (*Ladino* vs Indigenous). At the most basic level there is evidence of some heightened prevalence of stunting for Indigenous based on Table 1. To get a visual sense of the distributional issues we can compare the unconditional densities for the Indigenous and Ladino subpopulations (Figure 2). Comparing *Ladino* to Indigenous the quantiles shift left unequally with the smallest magnitude shift for the lowest quantile and the largest magnitude for the highest quantile. This reflects the move from a more dispersed to less dispersed distribution and is clearly evident in the figure. Notice that from the quantile shifts alone we cannot discern the implied changes in prevalence without assuming a functional form for both distributions. Because of the leftward shift, prevalence for stunting and severe stunting increase while prevalence of non-stunting decreases.

We need to move to a regression framework to control for compositional effects that might confound the simple unconditional distributional analysis of ethnicity. From the quantile regression results we see that after controlling for other covariates, the conditional shifts in quantiles retain the same ordering ($|\beta_{eth}(\tau = 0.26)| < |\beta_{eth}(\tau = 0.49)| < |\beta_{eth}(\tau = 0.91)|$); so the complex distributional features have persisted. Using simple linear regression for the same model would yield an ethnicity effect of -0.34 and would imply equivalent shifts of that amount for all quantiles.[15] As noted in section 2.3, the relatively smaller scale of the Indigenous distribution, controlling for other compositional effects, suggests that the subpopulation is more easily targeted than if there had been a leftward shift and an increase in scale.

The ordinal regression allows us to directly explore changes in prevalence that are not apparent from the quantile results. Using the two-stage estimator from section 2.2, yields the parameter estimates in Table 3 and post-processing of the bootstrap sample estimates using average predictive comparisons yields Figures 4 and 5. The expected changes in prevalence of severe stunting, stunting, and no stunting retain the same relative ordering as our analysis based on simple density comparisons, but the magnitudes change after controlling for covariates (see Figure 4). The change in prevalence of stunting, conditional on no severe

(1.95), Northeast (1.72), age (1.76), and water delivery (pipes inside 1.74, pipes outside 1.74, well 1.71, other 1.68). The test statistics for location-scale shift are Northeast (1.79), Southwest (1.85), water delivery (pipes inside 1.62, pipes outside 1.67, well 1.94, other 1.65), and education (primary 1.78, secondary or more 1.61). For the full models, the test statistics were 23.64 (location-shift) and 24.73 (location-scale) were below the critical values. Full tests results are available from the authors.

[15] Note that any predictions out of an OLS model that are used to then assess prevalence as $\frac{1}{n}\sum_{i=1}^{n} I(\hat{y}_i < -2)$ would be wrong since it imposes a symmetric distribution.

stunting is isolated in Figure 5, allowing us to assess the simple lower tail of a binary outcome rather than thinking about the complex set of changes that occur over multiple cut-points as a distribution shifts and changes shape. There is an increase in prevalence of almost 0.1 among indigenous even after controlling for other covariates. This is a result that can be easily adapted into the policy context in support of targeting. While both models provide complementary insights into the effect of ethnicity that are sensitive to changes in shapes of underlying distributions, the ordinal regression results give us direct insights into changes in prevalence that will be more easily digested and understood in policy settings.

The heteroskedasticity adjustment is evident in the changes in parameter estimates from the two-stage estimator that (Table 3) and unadjusted ordinal regression (Table A.5). The unadjusted $\tilde{\gamma}_{eth,(\mathbf{y}<-3)} = 0.11$ and the adjusted $\hat{\gamma}_{eth,(\mathbf{y}<-3)} = 0.26$ more than doubles; and the unadjusted $\tilde{\gamma}_{eth,(\mathbf{y}<-2)} = 0.63$ and adjusted $\hat{\gamma}_{eth,(\mathbf{y}<-2)} = 0.42$ while closer capture a clear adjustment effect. Those differences in parameter estimates also carry through to the average predictive comparisons. Also note that in the case where covariate estimates suggest only a location-shift effect (e.g. age or number of children), the unadjusted and adjusted ordinal regression parameter estimates are roughly equal.

The assessment of significant effects from the ordinal regression results can also benefit by referring to significance of the same effects in the quantile regression. As noted above, the heterskedastic-corrected estimates are successful in removing bias, but they are less efficient than the uncorrected estimates, and ordinal regression is already less efficient than quantile regression because of information loss in the transformation of a continuous variable to binary outcomes. However, since the ordinal regression effects are reflecting changes in quantiles through the distribution function, the information about significance in the quantile results can be used as an additional source of information in evaluating the results of the ordinal regression.

5. Discussion and Conclusions

In this paper we compared the results of two methods useful in evaluating correlates of child stunting. We first illustrated the theoretical and econometric links between two common statistical models: quantile and ordinal regressions. We then used child malnutrition in Guatemala as a case study, and demonstrate the insights that can be gained from a interpretation that reflects an integration of the insights drawn from each model.

At the most basic level of sign and significance of effects, the models provide us with results that are already known - ethnicity, sanitation and place of residence have notable impacts on malnutrition outcomes. Apart from the existence of numerous studies revealing these patterns, these relationships between malnutrition outcomes and various individual, household and parental characteristics are hinted at in basic descriptive tables presenting means and prevalences. Simple tabular analysis can be used to highlight differences in prevalence while regression models provide insightful nuance ensuring that the more complicated relationships are not masked by confounding or aggregation effects. Regression models therefore provide a frequently used strategy enabling analysts to *adjust* for characteristics not of prime interest. And while regression analysis represents a more complex analysis than cross-tabulations, even without extensive statistical training, policy-makers are fairly well versed in applying model results in a relevant way, primarily through the use of prevalence values. Based on this research, however, we suggest that employing a single regression modeling strategy may not indeed be adequate to fully capture the patterns underlying undernutrition in the developing world. We have demonstrated that by slightly increasing the complexity of the analytic strategy – joint consideration of quantile regression *and* ordinal regression models – the results provide a more complete representation of actual patterns and increase the policy utility of the interpretations.

We conclude that while the expected changes in prevalence based on estimates from ordinal regression are more easily explained in a policy setting, there are distributional features that enrich the interpretation that can only be isolated using quantile regression. Specifically, quantile regression allows analysts to identify possible changes in the scale or shape of the conditional distribution; information that may also prove useful for policy. We discussed the approach to policy framing in section 2.3 and return to policy considerations in our discussion of the results for Indigenous children. We noted that the quantile regression reveals a reduced scale, implying more uniform stunting among Indigenous and thus making the Indigenous a good target for intervention. The ordinal results complement this finding with an estimate of the difference in the prevalence of stunting among Indigenous and *Ladino* after controlling for compositional effects of the two populations. The prevalence difference provides a target for the magnitude of reduction that might be possible to achieve through targeted policy. We suggest that an analyst's understanding of the process under study, the policy utility, and a policy audience's ability to engage with the results are each improved if the two regression approaches are used in combination.

Finally, if quantile regression is warranted, we noted that ordinal regression results will be inconsistent. We proposed a two-stage estimator for ordinal regression using residuals from a first stage median regression to remove scaling effects in the index function of the ordinal regression model. The approach we describe in this paper could be useful in related research in which continuous data is grouped into discrete categories (this is frequently the case for analysis of anthropometry data).

There are a few remaining issues that we plan to take up in future research. While the estimation strategy proposed resolved a critical problem in making statistically valid comparisons between quantile and ordinal regression, there are other econometric issues that may present in other research settings. In this paper we proposed a strategy for an unfocused analysis to reveal basic patterns of covariation. If analysis instead focuses on a specific policy treatment effect, or more generally aims for causal interpretations of regression effects, the estimation strategy will also need to account for endogeneity. The instrumental variables estimator proposed by Chernozhukov and Hansen (2005, 2008) can be used to deal with endogeneity in quantile regression. For ordinal regression we suspect that the residual method proposed by (Terza et al., 2008) could be seemlessly imbedded in framework we proposed in this paper. More general issues of unobserved heterogeneity will need to be considered for panel data. This is an active research area for quantile regression. For ordinal regression we plan to investigate whether existing strategies for fixed or random effect estimators can be incorporated into the two-stage estimator introduced in this paper.

6. References

Abrevaya, J., 2001. The effects of demographics and maternal behavior on the distribution of birth outcomes. Empirical Economics 26 (1), 247–257.

Alderman, H., Hoddinott, J., Kinsey, B., 2006. Long term consequences of early childhood malnutrition. Oxford Economic Papers 58 (3), 450–474.

Aturupane, H., Deolalikar, A. B., Gunewardena, D., 2008. The determinants of child weight and height in Sri Lanka: A quantile regression approach. Working Papers RP2008/53, World Institute for Development Economic Research (UNU-WIDER).

Balk, D., Storeygard, A., Levy, M., Gaskell, J., Sharma, M., Flor, R., 2005. Child hunger in the developing world: An analysis of environmental and social correlates. Food Policy 30 (5-6), 584–611.

Bassolé, L., 2007. Child malnutrition in Senegal: Does access to public infrastructure really matter? A quantile regression analysis. Tech. rep., Université d'Auvergne.

Berg, A., Muscat, R., 1972. Nutrition and development: The view of the planner. American Journal of Clinical Nutrition 25 (2), 186.

Borooah, V. K., 2005. The height-for-age of Indian children. Economics & Human Biology 3 (1), 45–65.

Brennan, L., McDonald, J., Shlomowitz, R., 2004. Infant feeding practices and chronic child malnutrition in the Indian states of Karnataka and Uttar Pradesh. Economics & Human Biology 2 (1), 139–158.

Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. Econometrica 73 (1), 245–261.

Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: A robust inference approach. Journal of Econometrics 142 (1), 379–398.

Espindola, E., León, A., Martinez, R., Schejtman, A., 2005. Poverty, hunger and food security in Central America and Panama. Serie Políticas Sociales. Naciones Unidas/CEPAL.

Farrow, A., Larrea, C., Hyman, G., Lema, G., 2005. Exploring the spatial variation of food poverty in Ecuador. Food Policy 30 (5-6), 510–531.

Fenske, N., Kneib, T., Hothorn, T., 2009. Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. Tech. rep., Department of Statistics, University of Munich.

Gelman, A., Pardoe, I., 2007. Average predictive comparisons for models with nonlinearity, interactions, and variance components. Sociological Methodology 37 (1), 23–51.

Greene, W., 2003. Econometric Analysis. 5th edition. Prentice Hall.

Handcock, M. S., Morris, M., 1999. Relative Distribution Methods in the Social Sciences. Statistics for Social Science and Public Policy. Springer.

Hao, L., Naiman, D., 2007. Quantile Regression. Quantitative Applications in the Social Sciences. SAGE Publications.
Harrell, F., 2001. Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis. Springer Verlag.
INE, 2000. Encuesta Nacional de Condiciones de Vida. Tech. rep., Instituto Nacional de Estadistica de Guatemala.
Jamison, D., 1986. Child malnutrition and school performance in China. Journal of Development Economics 20 (2), 299–309.
Kandpal, E., McNamara, P., 2009. Determinants of nutritional outcomes of children in India: A quantile regression approach. In: 2009 Annual Meeting of the Agricultural and Applied Economics Association, July 26-28, 2009, Milwaukee, Wisconsin.
Koenker, R., 2005. Quantile Regression. Cambridge University Press.
Koenker, R., 2010. quantreg package, version 4.50.
Koenker, R., Bassett, Gilbert, J., 1978. Regression quantiles. Econometrica 46 (1), 33–50.
Koenker, R., Xiao, Z., 2002. Inference on the quantile regression process. Econometrica 70 (4), 1583–1612.
Larrea, C., Kawachi, I., 2005. Does economic inequality affect child malnutrition? The case of Ecuador. Social Science & Medicine 60 (1), 165–178.
Lee, J., Houser, R., Must, A., de Fulladolsa, P., Bermudez, O., 2010. Disentangling nutritional factors and household characteristics related to child stunting and maternal overweight in Guatemala. Economics & Human Biology 8 (2), 188–196.
Marini, A., Gragnolati, M., 2003. Malnutrition and Poverty in Guatemala. World Bank, Latin America and the Caribbean Region, Human Development Sector Unit.
Newey, W., Powell, J., 1990. Efficient estimation of linear and type i censored regression models under conditional quantile restrictions. Econometric Theory 6 (3), 295–317.
Pebley, A., Goldman, N., 1995. Social inequality and children's growth in Guatemala. Health transition review: the cultural, social, and behavioural determinants of health 5 (1), 1.
Sahn, D. E., Stifel, D., 2002. Robust comparisons of malnutrition in developing countries. American Journal of Agricultural Economics 84 (3), 716–735.
Seoane, N., Latham, M., 1971. Nutritional anthropometry in the identification of malnutrition in childhood. Journal of Tropical Pediatrics 17 (3), 98.
Stone, C., 1977. Consistent nonparametric regression. The Annals of Statistics 5 (4), 595–620.
Strauss, J., Thomas, D., 1998. Health, nutrition, and economic development. Journal of Economic Literature XXXVI, 766–817.
Sturm, R., Datar, A., 2005. Body mass index in elementary school children, metropolitan area food prices and food outlet density. Public Health 119 (12), 1059–1068.
Terza, J. V., Basu, A., Rathouz, P. J., 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. Journal of Health Economics 27 (3), 531–543.
WHO, 2004. The WHO Multicentre Growth Reference Study (MGRS). Tech. rep., World Health Organization.
WHO, 2010. World health statistics 2010. Tech. rep., World Health Organization.
Winshop, C., Mare, R., 1984. Regression models with ordinal variables. American Sociological Review 49 (4), 512–525.
Zhao, Q., 2001. Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form. Econometric Theory 17, 765–784.

Variable	Severely Malnourished HAZ<-3	Malnourished $-3 \leq$ HAZ <-2	Healthy HAZ $\geq$-2	Total
Observations	402	350	768	1,520
HAZ	-3.92	-2.47	-0.69	-2.04
Urban	0.27	0.28	0.48	0.37
Region: Metropolitan	0.20	0.15	0.25	0.21
Region: North	0.10	0.10	0.09	0.10
Region: Northeast	0.04	0.09	0.13	0.09
Region: Southeast	0.07	0.11	0.08	0.09
Region: Central	0.10	0.11	0.12	0.11
Region: Southwest	0.28	0.28	0.22	0.25
Region: Northwest	0.19	0.12	0.08	0.12
Region: Peten	0.02	0.04	0.05	0.04
Sex (male)	0.56	0.49	0.52	0.52
Age (months)	3.68	3.83	3.63	3.69
Parents' height	-0.36	-0.24	0.35	0.01
Indigenous	0.57	0.57	0.26	0.42
Water (surface)	0.11	0.10	0.07	0.09
Water (pipes in house)	0.38	0.49	0.59	0.51
Water (pipes on property)	0.17	0.13	0.13	0.14
Water (piped from well)	0.09	0.02	0.01	0.04
Water (other)	0.25	0.26	0.20	0.23
Breastfed 6 months	0.61	0.64	0.71	0.66
Mother's education (None)	0.58	0.43	0.31	0.42
Mother's education (Primary)	0.32	0.48	0.42	0.40
Mother's education (Secondary+)	0.10	0.08	0.27	0.18
Number of children	2.08	2.00	1.71	1.88
Bare floor	0.72	0.54	0.36	0.51
Poverty	0.75	0.75	0.48	0.62

Table 1: Weighted means and proportions

Variable	$\tau = 0.26$		$\tau = 0.49$		$\tau = 0.92$	
	β	SE	β	SE	β	SE
Intercept	-1.597**	(0.319)	-0.978**	(0.256)	1.262.	(0.657)
Urban	0.050	(0.111)	0.166.	(0.099)	0.464**	(0.170)
Region: North	0.080	(0.172)	0.287.	(0.169)	0.095	(0.368)
Region: Northeast	0.042	(0.177)	0.059	(0.176)	0.402	(0.334)
Region: Southeast	-0.306	(0.200)	-0.171	(0.168)	0.031	(0.319)
Region: Central	-0.267.	(0.151)	-0.104	(0.156)	-0.199	(0.251)
Region: Southwest	-0.480**	(0.163)	-0.155	(0.187)	0.369	(0.309)
Region: Northwest	-0.523**	(0.155)	-0.306*	(0.150)	-0.046	(0.354)
Region: Peten	0.234	(0.172)	0.133	(0.144)	-0.288	(0.293)
Sex (male)	-0.176*	(0.081)	-0.151.	(0.078)	-0.059	(0.153)
Age (months)	-0.095**	(0.034)	-0.119**	(0.029)	-0.225**	(0.062)
Parents' height	0.544**	(0.067)	0.555**	(0.058)	0.520**	(0.093)
Indigenous	-0.189.	(0.112)	-0.327**	(0.102)	-0.559**	(0.214)
Water (pipes in house)	0.017	(0.197)	0.160	(0.146)	-0.350	(0.506)
Water (pipes on property)	-0.330	(0.212)	-0.167	(0.175)	-0.800	(0.512)
Water (piped from well)	-0.723*	(0.326)	-0.492	(0.311)	-0.752	(0.612)
Water (other)	-0.144	(0.201)	0.049	(0.150)	-0.179	(0.486)
Breastfed 6 months	0.133	(0.093)	0.108	(0.089)	0.194	(0.152)
Mother's education (Primary)	0.236*	(0.106)	0.155	(0.102)	0.054	(0.185)
Mother's education (Secondary+)	0.419**	(0.139)	0.207	(0.143)	0.019	(0.224)
Number of children	-0.169**	(0.061)	-0.231**	(0.054)	-0.236*	(0.100)
Bare floor	-0.280**	(0.102)	-0.145.	(0.084)	-0.011	(0.169)
Poverty	-0.193	(0.118)	-0.241*	(0.098)	-0.186	(0.181)

Table 2: Quantile regression results. Standard errors are based on a pairs cluster bootstrap estimator. Symbols indicate results of tests for $H_o : \beta(\tau) = 0$; $\alpha \leq 0.01$ (**), $\alpha \leq 0.05$ (*), $\alpha \leq 0.1$ (.),$\alpha > 0.1$ ()

Variable	$\mathbf{y} < -3$		$\mathbf{y} < -2$	
	γ	SE	γ	SE
Intercept	-1.645**	(0.433)	-1.209**	(0.392)
Urban	-0.310.	(0.167)	-0.235.	(0.123)
Region: North	-0.360	(0.290)	-0.318	(0.235)
Region: Northeast	-0.360	(0.467)	0.051	(0.246)
Region: Southeast	-0.205	(0.324)	0.228	(0.236)
Region: Central	0.053	(0.266)	0.014	(0.212)
Region: Southwest	-0.012	(0.287)	0.134	(0.235)
Region: Northwest	0.181	(0.256)	0.324	(0.212)
Region: Peten	-0.562.	(0.342)	-0.053	(0.219)
Sex (male)	0.138	(0.117)	0.140	(0.109)
Age (months)	0.042	(0.047)	0.147**	(0.041)
Parents' height	-0.669**	(0.102)	-0.652**	(0.091)
Indigenous	0.256	(0.155)	0.424**	(0.134)
Water (pipes in house)	-0.114	(0.206)	-0.168	(0.233)
Water (pipes on property)	0.203	(0.266)	0.084	(0.280)
Water (piped from well)	1.077*	(0.524)	0.737	(1.043)
Water (other)	-0.175	(0.221)	-0.054	(0.238)
Breastfed 6 months	-0.271.	(0.132)	-0.081	(0.122)
Mother's education (Primary)	-0.354.	(0.156)	-0.144	(0.134)
Mother's education (Secondary+)	-0.876**	(0.402)	-0.319.	(0.197)
Number of children	0.233*	(0.089)	0.277**	(0.071)
Bare floor	0.297.	(0.142)	0.069	(0.113)
Poverty	0.292	(0.195)	0.242.	(0.125)

Table 3: Ordinal regression results. Standard errors are based on a pairs cluster bootstrap estimator.

Variable	Null Hypothesis: $\beta(\tau_i) = \beta(\tau_j)$		
	$\tau = (0.26, 0.49)$	$\tau = (0.49, 0.91)$	$\tau = (0.26, 0.91)$
Urban	0.203	0.029	0.007
Region: North	0.190	0.307	0.292
Region: Northeast	0.459	0.121	0.112
Region: Southeast	0.242	0.328	0.192
Region: Central	0.292	0.408	0.326
Region: Southwest	0.032	0.017	0.000
Region: Northwest	0.156	0.105	0.033
Region: Peten	0.439	0.208	0.264
Sex (male)	0.231	0.147	0.064
Age (months)	0.281	0.003	0.001
Parents' height	0.307	0.305	0.307
Indigenous	0.176	0.038	0.009
Water (pipes in house)	0.366	0.254	0.295
Water (pipes on property)	0.391	0.131	0.192
Water (piped from well)	0.406	0.466	0.506
Water (other)	0.253	0.418	0.446
Breastfed 6 months	0.372	0.337	0.335
Mother's education (Primary)	0.273	0.289	0.235
Mother's education (Secondary+)	0.113	0.288	0.112
Number of children	0.229	0.287	0.207
Bare floor	0.131	0.219	0.079
Poverty	0.297	0.253	0.319

Table 4: P-values from pairwise difference in slope tests.

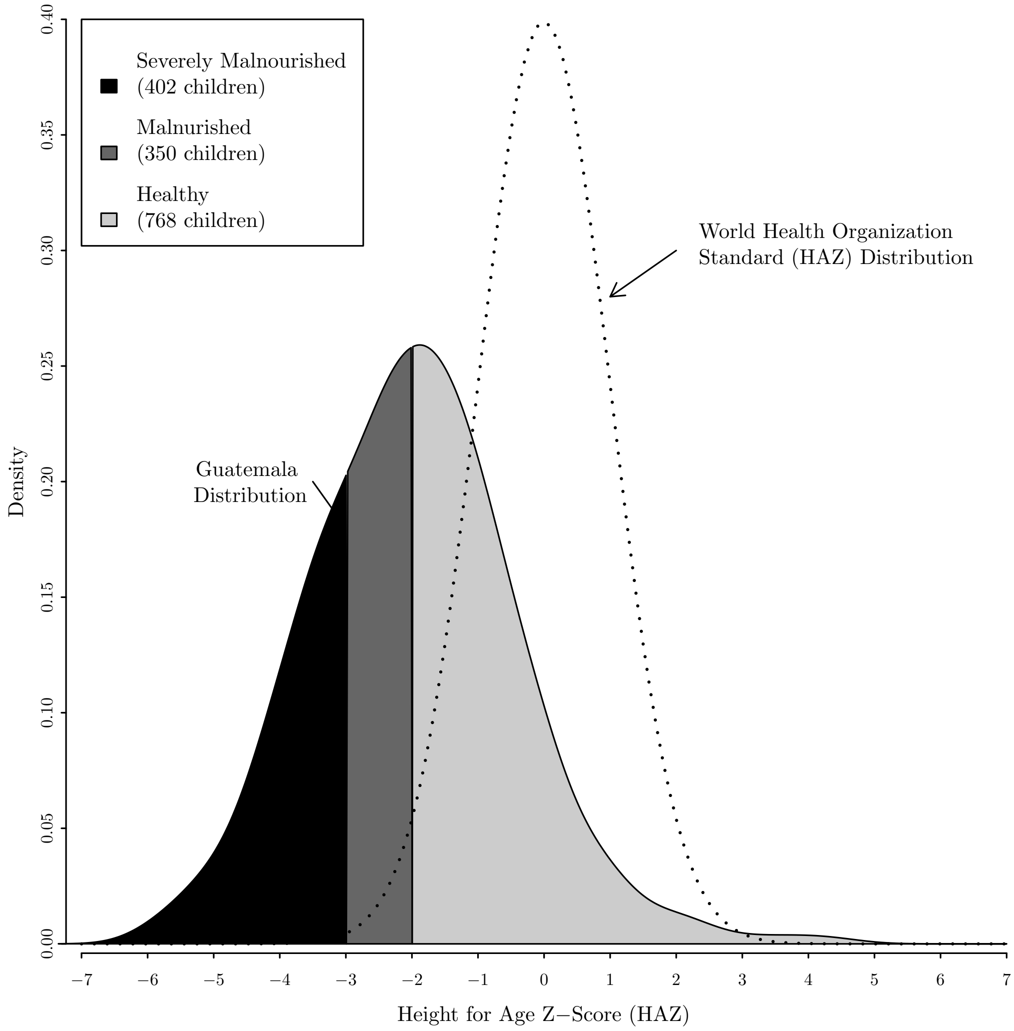


Figure 1: HAZ distribution for children aged 12 to 36 months

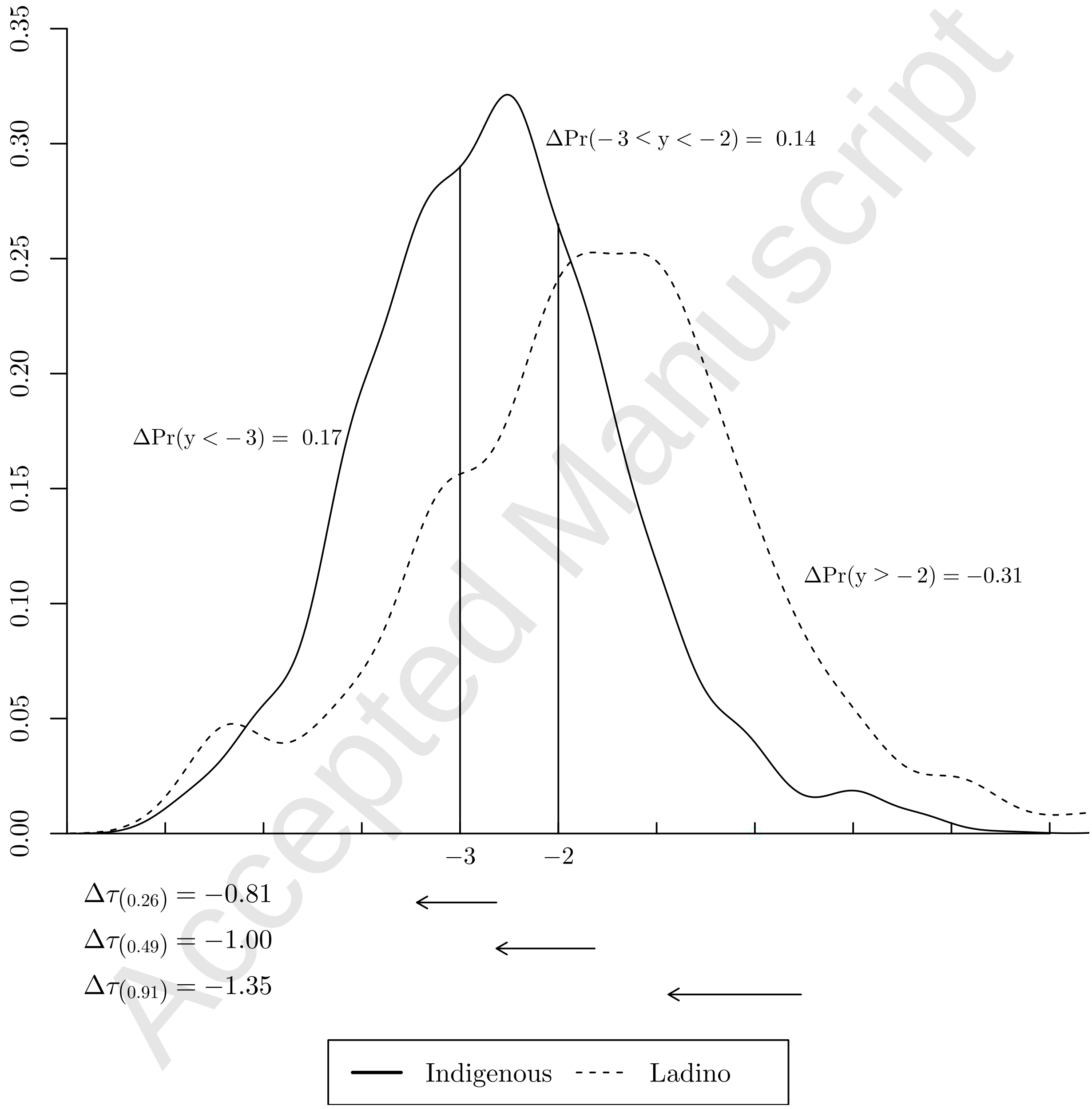


Figure 2: Unconditional densities comparing Indigenous to Ladino.

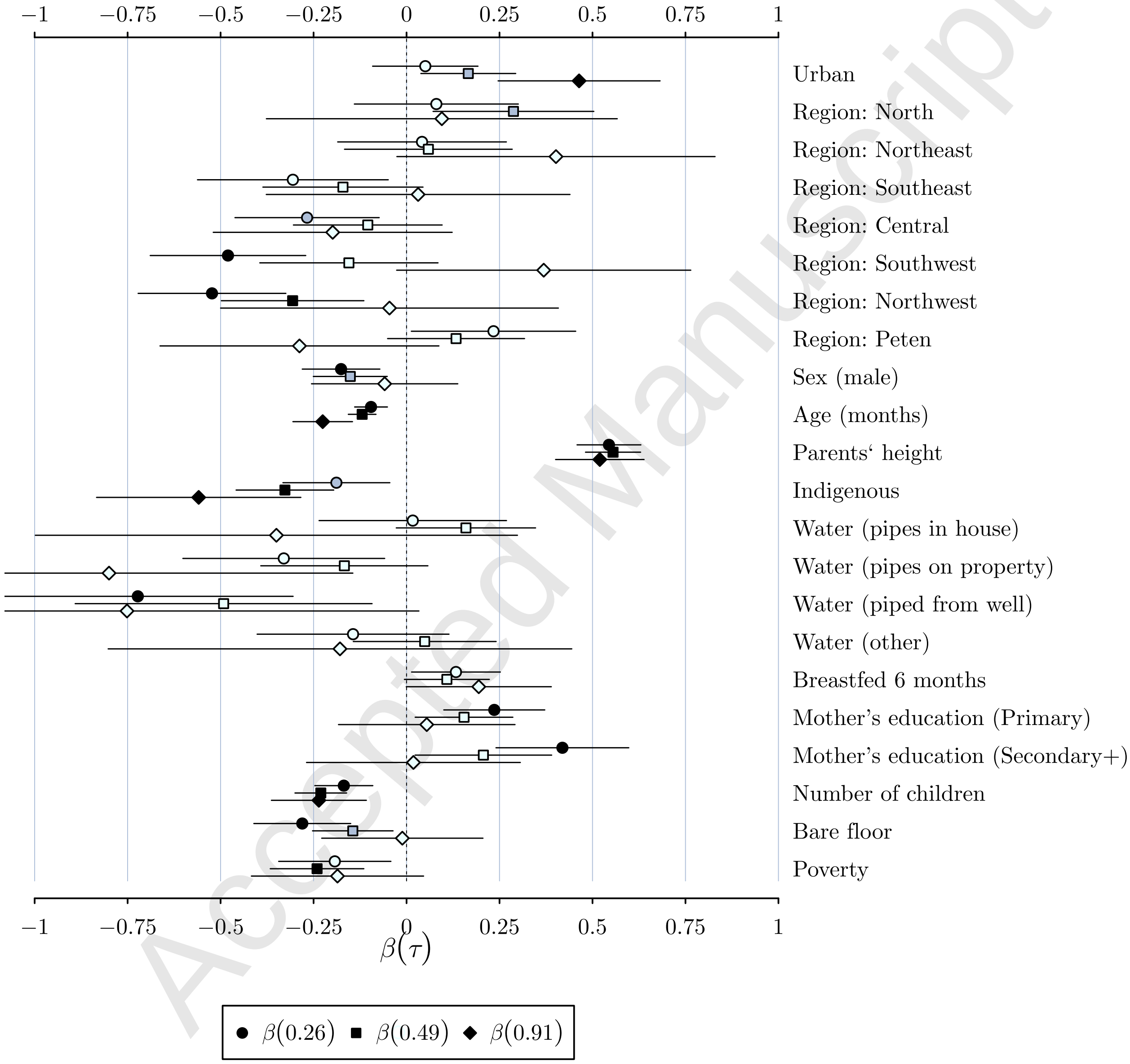


Figure 3: Quantile regression results. Standard errors are based on pairs cluster bootstrap. Shading indicates results of tests for $H_o : \beta(\tau) = 0$; $\alpha \leq 0.05$ (black), $\alpha \leq 0.1$ (gray), $\alpha > 0.1$ (white)

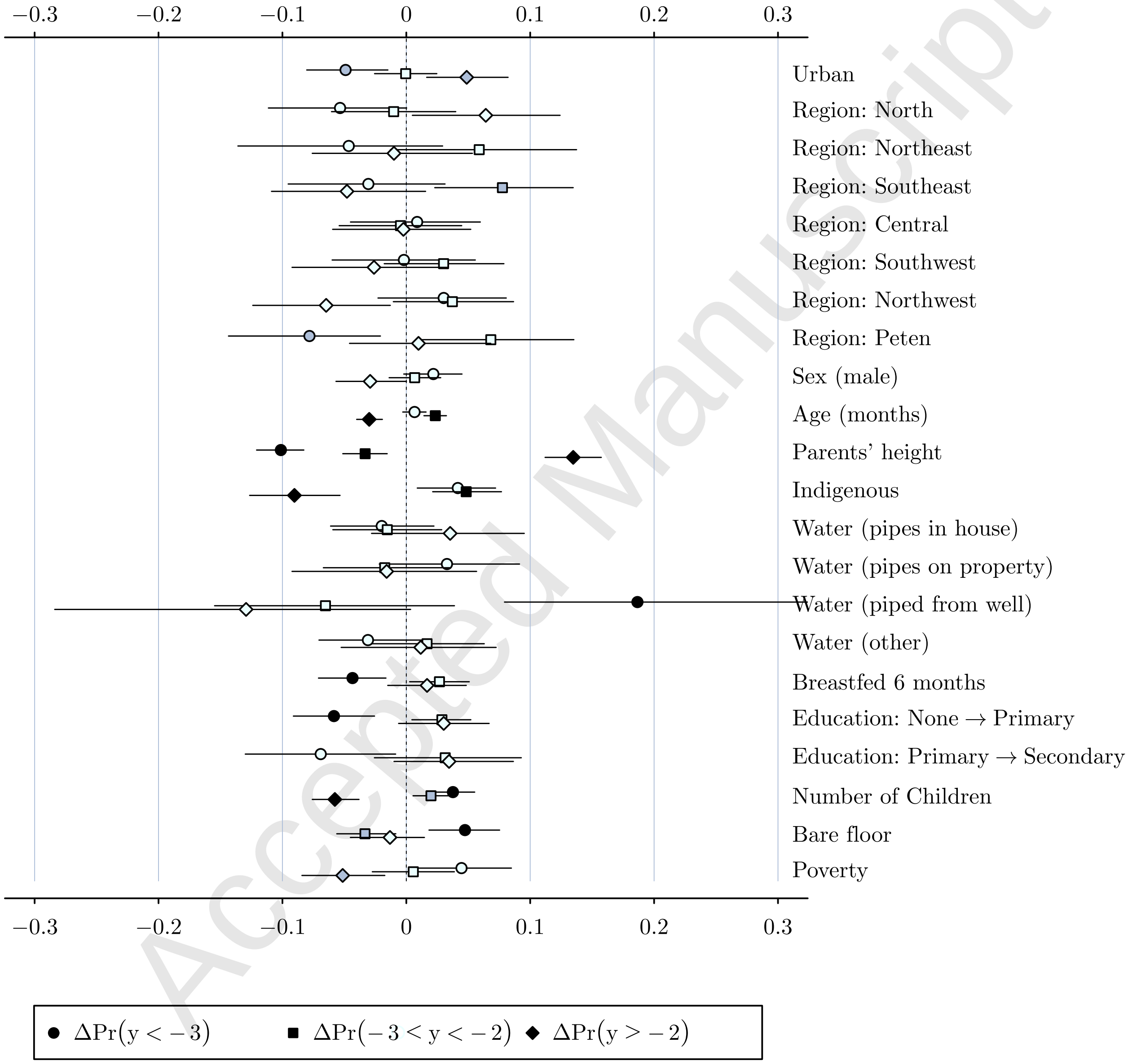


Figure 4: Ordinal regression results. Average Predictive Comparisons of unconditional probabilities. Shading indicates results of tests for $H_o : \beta(\tau) = 0$; $\alpha \leq 0.05$ (black), $\alpha \leq 0.1$ (gray), $\alpha > 0.1$ (white)

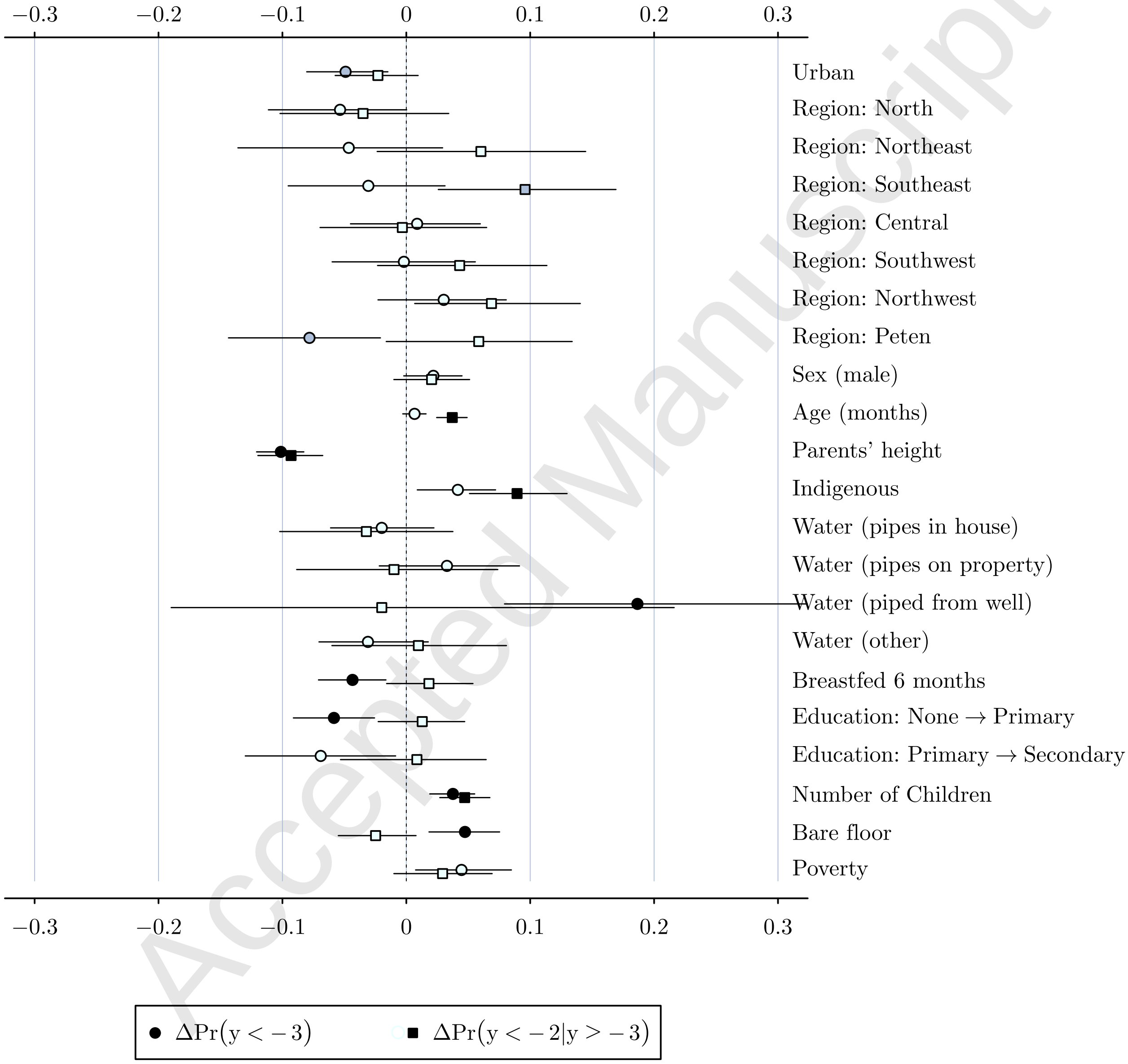


Figure 5: Ordinal regression results. Average Predictive Comparisons of continuation ratio probabilities. Shading indicates results of tests for $H_o : \beta(\tau) = 0$; $\alpha \leq 0.05$ (black), $\alpha \leq 0.1$ (gray), $\alpha > 0.1$ (white)

Appendices

Appendix A. Ordered logit parameters with no heteroskedasticity adjustment

Variable	$\mathbf{y} < -3$		$\mathbf{y} < -2$	
	γ	SE	γ	SE
Intercept	-1.594**	(0.519)	-1.504**	(0.480)
Urban	-0.336.	(0.205)	-0.324.	(0.159)
Region: North	-0.422	(0.414)	-0.465	(0.314)
Region: Northeast	-0.702	(0.459)	-0.134	(0.352)
Region: Southeast	-0.259	(0.406)	0.393	(0.310)
Region: Central	0.040	(0.363)	-0.002	(0.289)
Region: Southwest	0.162	(0.374)	0.204	(0.296)
Region: Northwest	0.187	(0.362)	0.204	(0.307)
Region: Peten	-0.943.	(0.394)	-0.211	(0.303)
Sex (male)	0.221.	(0.135)	0.161	(0.128)
Age (months)	0.037	(0.055)	0.179**	(0.053)
Parents' height	-0.751**	(0.115)	-0.926**	(0.107)
Indigenous	0.114	(0.176)	0.625**	(0.164)
Water (pipes in house)	-0.140	(0.257)	-0.200	(0.257)
Water (pipes on property)	0.396	(0.312)	0.180	(0.290)
Water (piped from well)	1.523**	(0.427)	1.081.	(0.639)
Water (other)	0.051	(0.266)	-0.034	(0.253)
Breastfed 6 months	-0.320.	(0.139)	-0.116	(0.136)
Mother's education (Primary)	-0.533**	(0.162)	-0.102	(0.147)
Mother's education (Secondary+)	-0.709*	(0.303)	-0.370	(0.253)
Number of children	0.249**	(0.092)	0.320**	(0.079)
Bare floor	0.417**	(0.163)	0.117	(0.147)
Poverty	0.113	(0.198)	0.246	(0.160)

Table A.5: Ordered logit parameters, no adjustment for heteroskedasticity