UNITED NATIONS UNIVERSITY
UNU-WIDER

This paper is a draft submission to the

WIDER Development Conference

Human capital and growth

6-7 June 2016 Helsinki, Finland

Catch-up Effect in Health Outcomes – Linear and Quantile Regression Estimates from Four Countries*

Subha Mani†
Jere R. Behrman‡
Andreas Georgiadis§
Fan Wang¶
Young Lives Determinants and Consequences of Child Growth Project Team‖

Do not cite without permission

Abstract

In many of the developing countries, even today, more than one-third of children under the age of 5 suffer from chronic nutritional deficiencies. Yet, little is known about the extent to which these children are able to recover from some of the long-term deficits in health outcomes caused by childhood undernourishment. To capture the association between nutritional status at young ages and subsequent health, we estimate a dynamic linear panel data model using data from the three waves of the Young Lives Study. We find that the catch-up coefficient in the linear dynamic panel data model varies between 0 and 0.23, where, Ethiopia and India exhibit perfect catch-up and Peru and Vietnam exhibit partial catch-up in height-for-age z scores. To further allow the catch-up coefficient to vary along the distribution of child anthropometrics, we also estimate a dynamic quantile regression model. We find that the null of homogenous catch-up effect along the distribution of height-for-age z scores can be easily rejected.

*This study is based on research funded by the Bill & Melinda Gates Foundation (Global Health Grant OPP1032713), Eunice Shriver Kennedy National Institute of Child Health and Development (Grant R01 HD070993) and Grand Challenges Canada (Grant 0072-03 to the Grantee The Trustees of the University of Pennsylvania). The data come from Young Lives, a 15-year survey investigating the changing nature of childhood poverty in Ethiopia, India (Andhra Pradesh), Peru and Vietnam (www.younglives.org.uk). Young Lives is core-funded by UK aid from the Department for International Development (DFID) and co-funded from 2010 to 2014 by the Netherlands Ministry of Foreign Affairs. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation, the Eunice Shriver Kennedy National Institute of Child Health and Development, Young Lives, DFID or other funders. The funders had no involvement in study design; in the collection, analysis and interpretation of data.

†Subha Mani, Department of Economics, Fordham University

‡Jere R. Behrman, Economics and Sociology Departments, University of Pennsylvania

§Andreas Georgiadis, Young Lives Study, Department of International Development, University of Oxford

¶Department of Economics, University of Houston

‖The Young Lives Determinants and Consequences of Child Growth Project team includes, in addition to co-authors of this paper, Benjamin T. Crookston, Santiago Cueto, Kirk Dearden, Le Thuc Duc, Patrice Engle, Javier Escobal, Lia Fernald, Shaik Galab, Nafisa Halim, Priscila Hermida, Elizabeth A. Lundeen, Mary Penny, Whitney Schott, Aryeh D. Stein, Tassew Woldehanna.

1 Introduction

Chronic malnutrition as measured by stunting among children under the age of 5 is associated with few grades of schooling, lower test scores and smaller stature as an adult. These factors further limit an individual's lifetime earnings and well-being [Stein et. al 2003, 2006, 2008; Hoddinott et. al 2008, 2010; Victora et. al 2008; Behrman et. al 2009; Maluccio et. al 2009]. However, observed catch-up in stature among children can minimize the permanent deficits of growth faltering. Existing evidence from dynamic linear panel data models depict that children are able to recover from some, on an average between one-third to one-fourth, but not all of the deficits in health caused by early nutritional deficiencies, concluding partial catch-up in health status [see Hoddinott and Kinsey 2001; Fedorov and Sahn 2005; Alderman, Hoddinott and Kinsey 2006; Mani 2011].

Policy makers are often most interested in identifying the extent to which catch-up is observed among children at the bottom of the nutritional distribution, that is, to what extent are poor nutritional outcomes observed today a reflection of the past? The dynamic linear panel data model currently used in the literature assumes that the catch-up coefficient has the same impact along the entire distribution of anthropometric outcomes. We test this assumption by estimating the dynamic quantile regression model, which allows the impact of lagged health (the catch-up coefficient), household characteristics, individual characteristics and community characteristics to vary along the entire distribution of health outcomes. In particular, we are able to test the assumption of homogenous catch-up effect along the distribution of height-for-age z score. Before identifying the varying distributional effects, we will first estimate the dynamic linear panel data model using the Arellano-Bond (1991) estimator where the coefficient estimate on the lagged dependent variable captures the extent of recovery – complete, partial, or none, from chronic malnourishment.

We will estimate path dependence in two measures of health - height-for-age z score (continuous measure) and stunting (discrete measure) using data from four countries collected as part of the Young Lives study.[1] The Young Lives Study follows children at three critical ages - 1, 5, and 8. We have a unique panel where children from four different countries and backgrounds - Ethiopia, India, Peru, and Vietnam are all followed at the same age during the course of the entire study. Stunting is a serious source of concern among policy makers in these countries, more than 30% of children measured at age 1 suffer from chronic nutritional deficiencies. The magnitude of path dependence observed in children's health will predict the extent to which these early life deficiencies will affect the child's life time well-being.

The Arellano-Bond estimator results in a catch-up coefficient that varies from between 0 and 0.23. India and Ethiopia both exhibit perfect catch-up, that is, the null of zero path dependence between current height-for-age z score (HAZ) and one-period lagged HAZ cannot be rejected at even the 10% significance level. Vietnam and Peru exhibit partial catch-up, that is, the null of zero path dependence between current HAZ and one-period lagged HAZ can be rejected at the 1% significance level. Partial catch-up effect indicates that malnutrition during childhood causes some but not significant growth retardation in future health and well-being. We find similar effects observed in stunting, there is

[1] Stunting is defined as a categorical variable which takes a value 1 if the child has height-for-age z score is less than -2 and 0 otherwise

significant amount of path dependence in stunting for children in Vietnam, followed by children in Peru and India. However, children in Ethiopia exhibit no path dependence in stunting depicting maximum potential for recovery. Note that Ethiopia also starts out with the highest levels of stunting making this catch-up possible.

To further allow for varying catch-up effects along the entire distribution of nutritional outcomes, we also estimate a dynamic quantile regression model treating the lagged dependent variable as endogenous. The coefficient estimate on the lagged dependent variable captures the extent of catch-up observed for individuals at different points in the anthropometric distribution. If history does not matter, then children at the bottom quantile of the nutritional distribution must have no association with their lagged health status. However, if indeed, factors during early childhood continue to affect their later well being then, we are likely to observe high levels of path dependence between current and lagged health, especially at the bottom quantile of the nutritional distribution. In particular, we will not be able to reject the null of no path dependence between current and lagged health for children in the bottom quantile of the anthropometric distribution.

Our quantile regression instrument variable (QR-IV) estimates suggest that children exhibit different levels of catch-up along the entire distribution of anthropometric outcomes, and this effect varies across countries. Vietnam exhibits somewhat higher levels of path dependence between current and lagged health at the lowest quantile but not at the top quantile of the anthropometric distribution. Peru exhibits almost similar level of partial catch-up along the entire distribution of HAZ. India and Ethiopia both exhibit small levels of path dependence at the bottom quantiles and higher dependence at the top quantiles. The varying degree of path dependence captures the need for differential policy implications for these countries. In particular, Vietnam starts out most well-developed and hence, children who suffer from chronic malnutrition in Vietnam tend to stay behind much longer than children who don't in other countries like Ethiopia and India. This suggests that inequality in HAZ in Vietnam is likely to sustain in the long-run.

This paper contributes to the existing literature in many ways. First, the paper brings out the extent to which there is path dependence in two measures of health outcomes – height-for-age z scores and stunting, using a unique panel data on children from four countries: Ethiopia, India, Peru, and Vietnam. This paper is the first to provide empirical evidence on catch-up effects in three (India, Peru and Vietnam) out of these four countries. Second, to our knowledge the paper is also the first in this literature to allow for the impact of the lagged dependent variable to vary along the entire distribution of anthropometric outcomes, testing the assumption of constant catch-up effects. In doing so, we treat the lagged dependent variable as endogenous.

The rest of the paper is organized as follows. Section 2 describes the data; section 3 outlines the empirical specification to be estimated in section 4. The empirical results are discussed in section 4, robustness analysis discussed in section 5 and concluding remarks follow in section 6.

2 Data

The data used in this paper comes from the Young Lives Study, a large panel study of children covering four countries and more than 12,000 children in two cohorts: about

8,000 children enrolled in 2002 at age 6-18 months and about 4,000 children enrolled in 2002 at age 8 years. The young lives panel survey subsequently follows these children in 2006 and 2009. This paper uses only data from the younger cohort, which was born at about the same time as the Millennium Development Goals (MDGs) were established; indeed, the Young Lives study was designed explicitly to track progress toward the MDGs. Complete details on the sampling strategy are available at http://www.younglives.org.uk. The panel nature of the data allows us to capture transition in anthropometrics outcomes during three critical periods – age 1, age 5, and age 8 of a child's life. The survey obtains detailed information on household demographic characteristics, assets, parental education, and height. The survey also includes a detailed community level questionnaire which records details on prices of important food items, prices of medicines, and community infrastructure.

Height-for-age z score from 2002 is available for 1946 children in Ethiopia, of whom we are unable to trace only 10.12% of these children in the second and third waves of the Young Lives Study. In India, out of 1922 children surveyed during the first wave, we are unable to follow 8.53% of the original sample. In Vietnam, out of 1990 children initially surveyed, we are unable to follow only 7.79% of the original children through the second and third waves of the Young Lives Study. In Peru, of the 2040 children initially surveyed in 2002, we are unable to trace only 9.80% of these children through the rest of the waves of the study. Overall, among all four countries - the attrition rates are quite similar, we are able to trace almost 90% of the initial sample during the follow-up surveys conducted in years 2006 and 2009 when children were of age 5 and 8 respectively.

Average baseline height-for-age-z score for attritors and non-attritors is described in appendix Table A1. We find no systematic difference in baseline HAZ for children in Ethiopia and Peru. In India, children who dropped out of the sample appear to have 0.24 standard deviation higher HAZ scores compared to children who stayed in the sample indicating negative selection in the panel sample. In Vietnam, children who dropped out of the sample appear to have 0.42 standard deviation lower HAZ in the baseline compared to non-attritors indicating positive selection into the panel sample. Hence, there is some concern for attrition related selection bias in the catch-up coefficient. We come back to this point later in the robustness section of the paper.

Table 1 depicts trends in mean height-for-age z scores and the percentage of children classified as stunted over the three waves of the Young Lives Study. Averages reported in Panel A, Table 1 suggest that there has been significant improvement in mean height-for-age z scores over time for children, almost 0.20 standard deviation improvement in height-for-age z scores occurs between the age of 1 and 8. Almost 30% of children at age 1 suffer from stunting and the worst incidence of stunting is observed in Ethiopia (41.5%). We find that in most countries the mean height-for-age z scores worsen between ages 1 and 5 and then improve between ages 5 and 8, except in Ethiopia which starts out worse and continues to improve over time. The percentage of children classified as stunted also increases between 2002 and 2006 and then declines between 2006 and 2009. Overall, both the incidence and intensity of chronic nutritional deficiencies seems to have declined for all four countries over time as measured by both the percentage of children classified as stunted and in terms of mean height-for-age z scores.

3 Empirical Specification

Following Fedorov and Sahn (2005) and Mani (2011), we estimate the following dynamic linear panel data model, where the coefficient estimate on the lagged dependent variable captures the extent of recovery from childhood malnutrition, also known as the 'catch-up' term. A coefficient of zero indicates 'complete catch-up', a coefficient of one on the lagged health status indicates 'no catch-up', and a coefficient between zero and one indicates 'partial catch-up' [Hoddinott and Kinsey 2001; Fedorov and Sahn 2005; Alderman, Hoddinott and Kinsey 2006; Mani 2011].

$$H_{it} = \beta_0 + \beta_1 H_{it-1} + \beta_2 Ethiopia * H_{it-1} + \beta_3 India * H_{it-1} + \beta_4 Peru * H_{it-1} + \sum_{j=1}^{R} \beta_j^X X_{jit} + \sum_{j=1}^{S} \beta_j^Z Z_{ji} + \epsilon_i + \epsilon_h + \epsilon_c + \epsilon_{it} \quad (1)$$

The outcome variable of interest in this paper is height-for-age z score (HAZ) and stunting (=1 if HAZ<=-2, 0 otherwise) – well established long-run indicators of individual health status. The variables Ethiopia, India and Peru are country specific dummies, where Vietnam is the omitted category. The catch-up coefficient, β_1 only captures the extent to which there is path dependence in health outcomes for children in Vietnam. The composite linear hypotheses – $\beta_1 + \beta_2$ gives us the catch-up coefficient for Ethiopia. Similarly, we can compute the catch-up coefficients for India and Peru as $\beta_1 + \beta_3$ and $\beta_1 + \beta_4$ respectively. The right hand side variables included in the regressions control for - age of the child in months, male dummy, male dummy interacted with age in months, asset index, rural dummy, availability of electricity, availability of health center, availability of hospital, price of sugar, price of oil, price of salt, price of deworming pills, price of amoxicillin, and price of oral rehydration medicines. The control variables are chosen to reflect standard models of determinants of human capital accumulation in child health [Thomas and Strauss, 1992, 2008].[2] Descriptive statistics on the outcome variable, regressors and instruments are described in Table 2.

There are four sources of unobservables in the dynamic specification (equation 1) - ϵ_i, ϵ_h, ϵ_c, and ϵ_{it}. ϵ_i captures the time-invariant individual-specific unobservables such as the child's inherent healthiness which affects his or her ability to absorb nutrients and fight diseases. ϵ_h captures all time-invariant household-specific unobservables reflecting parental preferences toward child health. ϵ_c captures all time-invariant community-specific unobservables like community endowments and political associations/connections. ϵ_{it} includes child specific time-varying unobservables such as expected future health shocks, current health shocks, and expected future prices of consumption goods and health inputs, all of which are unknown to the econometrician at date t.

The condition of zero correlation between the error term and explanatory variables may never be satisfied with the inclusion of the lagged dependent variable in the right hand side [Deaton (1997); Blundell and Bond (1998); Wooldridge (2002)]. Hence with H_{it-1} endogenous, standard OLS estimate of the catch-up coefficient is likely to be biased and inconsistent.

[2] Details on the construction of the asset index is available from the authors upon request.

We first-difference (FD) equation (1) to remove all time-invariant unobservables where the first-difference specification can be written as follows:

$$\delta H_{it} = \beta_1 \delta H_{it-1} + \beta_2 \delta Ethiopia * H_{it-1} + \beta_3 \delta India * H_{it-1} + \beta_4 \delta Peru * H_{it-1} + \sum_{j=1}^{R} \beta_j^X \delta X_{jit} + \delta \epsilon_{it} \quad (2)$$

OLS estimation applied to a first-difference specification magnifies the measurement error bias in lagged health [see Griliches and Hausman, 1986]. Therefore to address both omitted variables bias and measurement error bias, we follow the dynamic linear panel data estimation strategy proposed by Arellano Bond (1991) where, the first-differenced lagged dependent variable is instrumented with the two-period lagged dependent variable and two-period lagged Xs under the assumption of zero first-order and second-order serial correlation in the errors specified in the levels specification (1).

The standard linear specification described above assumes that the catch-up coefficient has the same impact across the entire distribution of anthropometric measurements, giving us the average partial effect of lagged health on current health. To allow and test for the possibility of varying distributional effects, we will also estimate the following dynamic panel data model using a Quantile-Regression framework as described below:

$$Q(\tau)_{hit} = \beta_0 + \beta_1(\tau) H_{it-1} + \beta_2(\tau) Ethiopia * H_{it-1} + \beta_3(\tau) India * H_{it-1} + \beta_4(\tau) Peru * H_{it-1} + \sum_{j=1}^{R} \beta_j^X(\tau) X_{jit} + \sum_{j=1}^{S} \beta_j^Z(\tau) Z_{ji} + \epsilon_{it} \quad (3)$$

where, the the impact of the lagged dependent variable, Xs and the Zs are allowed to depend upon the τ_{th} quantile of interest.

The main coefficient of interest is the parameter estimate on the lagged dependent variable which is allowed to vary along different quantiles and different countries. To address the endogeneity in the lagged dependent variable in equation (3), we will estimate the model using a Quantile-Regression-Instrumental-Variable (QR-IV) estimator proposed by Galvao (2011). The QR-IV estimator proposed by Galvao (2011) is similar to the Arellano-Bond estimator and uses two-period lagged dependent variable and two-period lagged Xs as instruments for the first-differenced lagged dependent variable in a quantile regression framework.

4 Results

4.1 Dynamic Linear Panel Data Estimates

The OLS estimates obtained from estimating the dynamic linear panel data specified in equation (1) is reported in column 1, Table 3. The OLS estimate on the catch-up coefficient is 0.70 (column 1, Table 3) for Vietnam and 0.61 (column 1, Table 3) for Peru, this indicates less than partial catch-up in attained height-for-age z score. We observe

almost partial catch-up in Ethiopia (0.42) and India (0.50). The OLS estimates though are likely to be biased and inconsistent as it suffers from both omitted variable bias and measurement error bias as discussed in the earlier sections of the paper.

Equation (2) is estimated using the Arellano-Bond (1991) estimator where the first-differenced catch-up coefficient and its interaction with the country dummies is instrumented with two-period lagged height-for-age z score, two-period lagged height-for-age z scores interacted with the country specific dummies, two-period lagged weight-for-age z score, two-period lagged weight-for-age z score and its interaction with the country specific dummies, two-period lagged availability of hospital and its interaction with country specific dummies. The FD-GMM estimates following the Arellano-Bond type estimator is reported in column 2, Table 3, which produces a catch-up coefficient estimate of 0.22 for Vietnam, 0.23 for Peru, -0.01 for Ethiopia, and 0.017 for India. We find that the null of perfect catch-up or no path dependence cannot be rejected at even the 10% significance level for children in Ethiopia and India, indicating no path dependence in health outcomes for children in these countries. We find evidence in support of more than partial catch-up among children in Vietnam and Peru, we can easily reject the null of no path dependence in health outcomes at the 1% significance level.

The coefficient on the catch-up term from the first-difference GMM/Arellano-Bond specification indicates larger catch-up effects compared to the coefficient estimate reported in the OLS specification, suggesting an upward bias in the OLS parameter estimate of the catch-up term. The catch-up coefficient obtained from following a first-difference GMM strategy provides us with our preferred estimate on the catch-up term as it addresses both omitted variables bias (via first-differencing) and measurement error bias (via instrumental-variable techniques) in data.

Similar evidence on path dependence in health outcomes is also observed for stunting reported in both the OLS estimates reported in columns 3 and the preferred first-difference GMM/Arellano-Bond estimates reported in column 4, Table 3. Where, the OLS estimates reported in column 3, Table 3 are biased upwards relative to the estimates reported in column 4, Table 3. The preferred first-difference GMM/Arellano-Bond estimates for stunting reported in column 4 indicate that a child who is stunted in the previous period in Vietnam is likely to be 38 percentage points more likely to be stunted today compared to a child who is not stunted in the previous period. Similarly, a child who is stunted in the previous period in Peru is 19 percentage points more likely to be stunted today compared to his/her counterpart who was not stunted in the last period. We also observe path dependence in stunting for children in India, where, a child who is stunted in the last period is 14 percentage points more likely to be stunted today compared to his/her counterpart in the last period. We are not able to find any path dependence in stunting for children in Ethiopia. We find substantial catch-up in health outcomes for children in Ethiopia followed by children in India, Peru and Vietnam.

There is limited variation over time in the community characteristics due to which most of the community variables are insignificant in the preferred first-difference specification.

4.2 Quantile Regression Estimates

To allow for catch-up effect to vary along the entire distribution of nutritional outcomes, we also estimate a dynamic quantile regression instrument variable model treating the

lagged dependent variable as endogenous. Equation (3) is estimated using the Quantile-Regression Instrument-Variable estimator proposed by Galvao (2011) where the first-differenced catch-up coefficient and its interaction with the country dummies is instrumented with two-period lagged height-for-age z score, two-period lagged height-for-age z score interacted with the country specific dummies, two-period lagged dummy capturing the incidence of diarrhea during the last 24 hours, and the interaction of this variable with the country specific dummies. The coefficient estimate on the lagged dependent variable captures the extent of catch-up observed for individuals at both the bottom and top quantiles of the anthropometric distribution. If history does not matter, then children at the bottom quantile of the nutritional distribution must have no association with their lagged health. However, if indeed, factors during early childhood continue to affect their later well being then, we will observe high levels of path dependence between current and lagged health, especially at the bottom quantile of the nutritional distribution. In particular, we will not be able to reject the null of no path dependence between current and lagged health for children in the bottom quantile of the anthropometric distribution.

The preferred instrument variable quantile regression estimates for the dynamic panel data model are reported in columns 1-4, Table 4, capturing the catch-up effects along the following four quantiles - τ=0.10 τ=0.25, τ=0.75, and τ=0.90. The instrument variable quantile regression estimates reported in Table 4 indicate that the null of homogenous catch-up coefficient can be rejected at the 1% significance level. We find that children exhibit different levels of catch-up along the distribution of anthropometric outcomes, and this effect varies across countries. Vietnam exhibit somewhat higher levels of path dependence between current and lagged health at the lowest quantile but not at the top quantile of the anthropometric distribution. India, Ethiopia, and Peru all exhibit small levels of path dependence at the bottom quantiles and higher dependence at the top quantiles.

5 Robustness

5.1 Instrument Validity

In the presence of weak correlation between the instruments and the endogenous regressors, the IV estimates reported here are likely to suffer from a larger bias and inconsistency compared to the bias obtained on the OLS parameter estimate (Murray, 2006). Recent work on weak IVs suggests the use of the Kleibergen-Paap Wald rk F statistic is robust to the presence of heteroskedasticity, autocorrelation and clustering (Kleibergen and Paap, 2006). The Kleibergen-Paap Wald rk F reported in columns 2 and 4 of Table 3 are greater than 10, satisfying the usual criteria for instrument validity, that is, our excluded IVs are strongly correlated with the endogenous regressor (Staiger and Stock, 1997).

In addition to the test of strong correlation between the endogenous regressor and the instrument, it must also be the case that the instrument is uncorrelated with the error term in the second stage regression. The Hansen J statistic (1996) reported in columns 2 and 4 of Table 3 suggest that we cannot reject the null of instrument validity for these specifications. The coefficient estimate on the Hansen J statistic and the first-stage F test statistic on the excluded instruments are all appended in Table 3.

Arellano and Bond (1991) stress that in their estimator, using a twice lagged depen-

dent variable (here H_{it-2}) as an instrument for first-differenced lagged dependent variable (H_{it-1}) is valid only if E (ϵ_{it}, H_{it-2}) = 0, that is, the errors in the levels specification are serially uncorrelated over time. To test for second-order serial correlation in the levels residuals, Arellano and Bond (1991, pp. 282) suggest using an m2 statistic. However, this requires a minimum of five rounds of data and the Young Lives Study has only three rounds of data. Instead, we use, the C statistic also known as the GMM distance or difference-in-Sargan statistic to test for the no serial correlation assumption in the levels specification (Blundell and Bond, 2000 and Hayashi, 2000).

The C-statistic tests of the serial correlation/exogeneity of the two-period lagged height-for-age z score and the interaction of this variable with the country dummies is reported in columns 2 and 4, Table 3. At the 10% significance level, we do not reject the null that the two-period lagged dependent variable and its interaction the country dummies is a valid instrument, that is, we cannot reject the null of no first-order and second-order serial correlation in the errors in levels specification.

The two conditions of instrument relevance discussed in this section provide additional support for the reliability of the preferred estimates obtained using the Arellano-Bond/first-difference GMM estimator as our preferred strategy.

5.2 Attrition

To determine the extent to which endogenous observables create attrition bias, we also estimate a linear probability model of sample attrition, where the dependent variable is defined as attrition which takes a value 1 if the child can be followed through the 2002, 2006, and 2009 waves of the Young Lives Study and 0 otherwise. The regression results on sample attrition are reported separately for each of the four countries in columns 1-4 in Appendix Table A2. In columns 1-4, attrition is regressed upon HAZ score from 2002 and baseline characteristics from 2002 which include age, male dummy, age interacted with the male dummy and the full set of community dummies to control for community specific time-invariant unobservables. The attrition regression results outlined in columns 1-4, Table A2 indicate that attrition is unrelated to the endogenous observable, HAZ score. Hence, our preferred estimates of the dynamic panel data model are not likely to be confounded by attrition bias.

The coefficient estimates on the catch-up term is also robust to sample attrition. Attrition can be a problem only if; either the observable or unobservables that result in attrition are correlated with the error term in the specification of interest (Fitzgerald et al. 1998). Individual level attrition is also not a real concern in this paper, given the estimation strategy adopted here. First-differencing removes all potential sources of unobservables like the childs genetic endowments which are likely to be correlated with potential observables or unobservables that result in attrition. In the presence of first-differencing, the only possible remaining source of attrition is that arising from the presence of random health shocks, such as infectious diseases. These health shocks are likely to be uncorrelated with the health shocks in subsequent periods. Hence, attrition arising from the presence of random, time-varying health shocks is not likely to contaminate the parameter estimate on the lagged dependent variable.

6 Conclusion

In this paper we estimate - (a) a linear dynamic panel data model to capture the extent of catch-up in health status among children, and (b) a dynamic quantile regression model, which allows the impact of the catch-up coefficient to vary along the entire distribution of anthropometric outcomes. Both models are estimated to account for endogeneity in the lagged dependent variable.

To address these questions, we will use panel data from four countries collected as part of the Young Lives study, following children at three critical ages - 1, 5, and 8 years. We have a unique panel where children from four different countries: Ethiopia, India, Peru, and Vietnam are all followed at the same age during the course of the entire study.

The FD-GMM estimator used to compute the coefficient estimate on the lagged dependent variable in the linear dynamic panel data model results in a catch-up coefficient that varies from between 0 and 0.23. India and Ethiopia both exhibit perfect catch-up, that is, the null of zero path dependence between current height-for-age z score (HAZ) and lagged HAZ cannot be rejected at even the 10% significance level. Vietnam and Peru exhibit partial catch-up, that is, the null of zero path dependence between current HAZ and lagged HAZ can be rejected at the 1% significance level. Partial catch-up effects indicate that malnutrition during childhood causes some but not significant growth retardation in future health and well-being. Since, Ethiopia and India both start out with very high levels of chronic nutritional deficiencies, there is maximum potential for recovery among children in these countries.

To further allow for the catch-up coefficient to vary along the entire distribution of nutritional outcomes, we also estimate a dynamic quantile regression instrument variable model treating the lagged dependent variable as endogenous. The coefficient estimate on the lagged dependent variable captures the extent of catch-up observed for individuals at both the bottom and top quantiles of the anthropometric distribution. If history does not matter, then children at the bottom quantile of the nutritional distribution must have no association with their lagged health status. However, if indeed, factors during early childhood continue to affect their later well being then, we will observe high levels of path dependence between current and lagged health, especially at the bottom quantile of the nutritional distribution. In particular, we will not be able to reject the null of no path dependence between current and lagged health for children in the bottom quantile of the anthropometric distribution.

We find that children exhibit different levels of catch-up along the distribution of anthropometric outcomes, and this effect varies across countries. Vietnam exhibit somewhat higher levels of path dependence between current and lagged health at the lowest quantile but not at the top quantile of the anthropometric distribution. India, Ethiopia, and Peru all exhibit small levels of path dependence at the bottom quantiles and higher dependence at the top quantiles. We reject the null of homogenous catch-up effect along the distribution of height-for-age z score for all countries, except Peru.

This paper contributes to the existing literature in many ways. First, the paper identifies the extent to which there is path dependence in height-for-age z score using data from a panel study on young children in four countries - Ethiopia, India, Peru, and Vietnam. This paper is the first to provide empirical evidence on catch-up effects in three of these four countries. Second, to our knowledge the paper is the first in this literature to allow

for the impact of the lagged dependent variable to vary across the entire distribution of anthropometric outcomes, testing the assumption of constant catch-up effects.

References

Alderman, H., Hoddinott, J., and Kinsey, W. 2006. Long-term consequences of early childhood malnutrition. *Oxford Economic Papers*, 58(3): 450-474.

Arellano, M. and Bond, S. 1991. Some tests of specification for panel data: Mote carlo evidence and an application of employment equations. *Review of Economic Studies*, 58: 277-297.

Behrman, Jere R., Maria Cecilia Calderon, Samuel Preston, John Hoddinott, Reynaldo Martorell and Aryeh D. Stein, 2009, Nutritional Supplementation of Girls Influences the Growth of their Children: Prospective Study in Guatemala, *American Journal of Clinical Nutrition* 90 (November 2009), 1372-1379.

Blundell, R. and S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87: 115-143.

Blundell, R., Bond, S., and Windmeijer, F. 2000. Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimators. Working Paper 00/12, *The Institute for Fiscal Studies.*

Blundell, R. and MaCurdy, T. 1999. Labor supply: A review of alternative approaches. *Handbook of Labor Economics*, III: 1560-169.

Blundell, R. 2005. Dynamic panel data methods. *Micro econometric lecture notes.*

Fedorov, L. and Sahn, D. 2005. Socioeconomic determinants of children's health in Russia: A longitudinal survey. *Economic Development and Cultural Change*, pp 479-500.

Galvao, A.F. 2011. Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, pp 142-157.

Hoddinott, J. and Kinsey, W. 2001. Child growth in the time of drought. *Oxford Bulletin of Economics and Statistics*, 63(4): 409-436.

Hoddinott, John, John A. Maluccio, Jere R Behrman, Rafael Flores and Reynaldo Martorell, 2008, The Impact of Nutrition During Early Childhood on Income, Hours Worked, and Wages of Guatemalan Adults *Lancet* 371 (February), 411-416.

Hoddinott, John, John Maluccio, Jere R. Behrman, Reynaldo Martorell, Paul Melgar, Agnes R. Quisumbing, Manuel Ramirez-Zea, Aryeh D. Stein, and Kathryn M. Yount, (2010). "The consequences of early childhood growth failure over the life course." Washington, DC: International Food Policy Research Institute.

Kleibergen, F. and R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133 (1): 97-126.

Maluccio, John A., John Hoddinott, Jere R. Behrman, Agnes Quisumbing, Reynaldo Martorell and Aryeh D. Stein, 2009,The Impact of Nutrition During Early Childhood on Education among Guatemalan Adults, *Economic Journal* 119 (April), 734763.

Mani, S. 2011. Is there Complete, Partial, or No Recovery from Childhood Malnutrition Empirical Evidence from Indonesia. *Oxford Bulletin of Economics and Statistics* doi: 10.1111/j.1468-0084.2011.00670.x

Mani, S., J. Hoddinott, J. Strauss. 2012. Long-term impact of investments in early schooling Empirical evidence from rural Ethiopia. *Journal of Development Economics.* Doi:10.1016/j.jdeveco.2012.03.002

Murray, M. P. 2006. Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, 20(4): 111-132.

Staiger, D. and Stock, J. H. 1997. Instrumental variables regression with weak instruments. *Econometrica*, 65(May): 557-86.

Stein, Aryeh D., Huiman X. Barnhart, Morgan Hickey, Usha Ramakrishman, Dirk G. Schroeder and Reynaldo Martorell, 2003, Prospective Study of Protein-Energy Supplementation Early in Life and of Growth in the Subsequent Generation in Guatemala, American Journal of Clinical Nutrition 78: 162-7.

Stein, Aryeh D., Meng Wang, Ann DiGirolamo, Ruben Grajeda, Usha Ramakrishnan, Manuel Ramirez-Zea, Kathyrn Yount and Reynaldo Martorell, 2008, Nutritional Supplementation in Early Childhood, Schooling and Intellectual Functioning in Adulthood: A Prospective Study in Guatemala, *Archives of Pediatric and Adolescent Medicine* 162:7 (July)., 612-618.

Stein, Aryeh D., Meng Wang, Manuel Ramirez-Zea, Rafael Flores, Ruben Grajeda, Paul Melgar, Usha Ramakrishnan, and Reynaldo Martorell,2006, Exposure to a Nutrition Supplementation Intervention in Early Childhood and Risk Factors for Cardiovascular Disease in Adulthood: Evidence from Guatemala, *American Journal of Epidemiology*

Strauss, J. and Thomas, D. 1995. Human resources: empirical modeling of household and family decisions. *Handbook of Development Economics*, vol. 3, edited by Jere R. Behrman and T.N. Srinivasan.

Strauss, J. and Thomas, D. 1998. Health, Nutrition and Economic Development, *Journal of Economic Literature*, 36(2): 766-817.

Strauss, J. and Thomas, D. 2008. Health over the life course, forthcoming in the *Handbook of Development Economics*, vol. 4, edited by T. Paul Schultz and John Strauss eds., Amsterdam: North Holland Press.

Victora, Cesar G , Linda Adair, Caroline Fall, Pedro C Hallal, Reynaldo Martorell, Linda Richter, Harshpal Singh Sachdev, on behalf of the Maternal and Child Undernutrition Study Group, 2008, Undernutrition 2: Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital, The *Lancet* 371 (Issue 9609), 340-357.

Table 1: Summary statistics – Height-for-age z-score

Panel A: Pooled sample - Ethiopia, India, Peru and Vietnam			
Years	% HAZ <=-2	Mean	Mean difference
2002	30.1	-1.33	-0.16*** (2006-2002)
		(0.016)	(0.02)
2006	31.18	-1.49	0.26*** (2009-2006)
		(0.012)	(0.017)
2009	22.74	-1.23	0.10*** (2009-2002)
		(0.012)	(0.02)
Panel B: Ethiopia			
Years	% HAZ <=-2	Mean	Mean difference
2002	41.5	-1.58	0.10** (2006-2002)
		(0.04)	(0.04)
2006	31.27	-1.48	0.26*** (2009-2006)
		(0.02)	(0.03)
2009	21.49	-1.22	0.36*** (2009-2002)
		(0.02)	(0.04)
Panel C: India			
Years	% HAZ <=-2	Mean	Mean difference
2002	31	-1.35	-0.30*** (2006-2002)
		(0.03)	(0.04)
2006	35.89	-1.65	0.20*** (2009-2006)
		(0.02)	(0.03)
2009	29.47	-1.45	-0.10*** (2009-2002)
		(0.02)	(0.04)
Panel D: Peru			
Years	% HAZ <=-2	Mean	Mean difference
2002	28.15	-1.30	-0.23*** (2006-2002)
		(0.02)	(0.03)
2006	32.77	-1.53	0.38*** (2009-2006)
		(0.02)	(0.03)
2009	20.27	-1.15	0.15*** (2009-2002)
		(0.02)	(0.03)
Panel E: Vietnam			
Years	% HAZ <=-2	Mean	Mean difference
2002	20.27	-1.09	-0.24*** (2006-2002)
		(0.02)	(0.03)
2006	24.85	-1.33	0.23*** (2009-2006)
		(0.02)	(0.03)
2009	19.72	-1.10	0.01 (2009-2002)
		(0.02)	(0.04)

Notes:

Source: Young Lives Study - 2002, 2006, and 2009

Standard errors reported in parentheses

*** significant at 1%, ** significant at 5%, * significant at 10%

Pooled sample = 21738; Ethiopia =5247; India = 5466; Peru = 5520; Vietnam =5505

Table 2: Summary statistics

Variables	Mean	Std. dev
Height-for-age z-score (HAZ)	-1.35	1.18
Weight-for-age z-score (WAZ)	-1.15	1.23
Age in months	57.45	34.89
Male (=1 if male, 0 for female)	0.52	0.49
Rural (=1 if rural, 0 otherwise)	0.63	0.48
Mother's height in cm	153.02	6.63
Caregiver's schooling	5.03	4.55
Asset index	0.0012	2.08
Electricity (=1 if available, 0 otherwise)	0.76	0.42
Hospital (=1 if available, 0 otherwise)	0.47	0.49
Health center (=1 if available, 0 otherwise)	0.86	0.34
Price of oil (in 2005 USD)	1.19	0.32
Price of sugar (in 2005 USD)	0.58	0.14
Price of salt (in 2005 USD)	0.16	0.08
Price of deworming medicines (in 2005 USD)	0.32	0.56
Price of oral rehydration medicines (in 2005 USD)	0.17	0.11
Price of Amoxycillin (in 2005 USD)	2.03	4.41
Sample size	21738	

Notes:

Source: Young Lives Study - 2002, 2006, and 2009

Table 3: Dynamic Linear Panel Data Estimates

	HAZ (1) OLS	HAZ (2) FD-GMM	Stunting (3) OLS	Stunting (4) FD-GMM
Lagged outcome	0.70***	0.22***	0.55***	0.38***
	(0.013)	(0.07)	(0.019)	(0.10)
Ethiopia*lagged outcome	-0.27***	-0.24***	-0.17***	-0.32***
	(0.02)	(0.08)	(0.03)	(0.11)
India*lagged outcome	-0.19***	-0.21**	-0.010	-0.28**
	(0.02)	(0.08)	(0.02)	(0.11)
Peru*lagged outcome	-0.08***	0.006	-0.078***	-0.18
	(0.017)	(0.11)	(0.03)	(0.11)
Male dummy	0.12***		-0.022*	
	(0.02)		(0.012)	
Lagged age in months	0.008***	-0.002	-0.002***	0.006*
	(0.0005)	(0.008)	(0.0002)	(0.003)
Lagaged age in months*	-0.003***	-0.0011***	0.0005***	
male dummy	(0.0004)	(0.0004)	(0.0002)	(0.0001)
Rural dummy	-0.10***	-0.19***	0.022*	0.11****
	(0.03)	(0.06)	(0.013)	(0.02)
Asset index	0.064***	0.015**	-0.024***	-0.003
	(0.006)	(0.007)	(0.003)	(0.004)
Electricity	-0.05	0.21**	0.013	-0.07
	(0.04)	(0.09)	(0.019)	(0.05)
Hospital	0.07**	0.07	0.013	-0.04
	(0.03)	(0.08)	(0.012)	(0.04)
Health center	0.046	-0.03	-0.02	0.009
	(0.03)	(0.09)	(0.017)	(0.08)
Price of oil	0.07	0.05	-0.02	-0.02
	(0.08)	(0.05)	0.018	(0.02)
Price of sugar	-0.18	-0.10	-0.04	0.07
	(0.19)	(0.14)	(0.04)	(0.07)
Price of salt	0.08	0.009	0.05	0.10
	(0.18)	(0.21)	(0.07)	(0.11)
Price of deworming pills	-0.03	-0.16	0.009	0.05
	(0.13)	(0.13)	(0.005)	(0.07)
Price of amoxicillin	-0.14***	-0.14**	0.056*	0.05
	(0.05)	(0.06)	(0.03)	(0.04)
Price of oral	0.52***	0.25**	-0.015	-0.17**
rehydration	(0.13)	(0.13)	(0.05)	(0.07)
Kleibergen-Paap F statistic		25.17		20.30
Hansen J statistic		11.07		11.92
		(0.43)		(0.15)
C statistic		5.98		7.49
		(0.20)		(0.12)
Catch-up coefficient				
Ethiopia	0.42***	-0.01	0.37***	0.05
	(0.017)	(0.02)	(0.02)	(0.037)
India	0.50***	0.017	0.54***	0.14***
	(0.019)	(0.02)	(0.02)	(0.05)
Peru	0.61***	0.23***	0.47***	0.19***
	(0.012)	(0.07)	(0.02)	(0.05)
Vietnam	0.70***	0.22***	0.55***	0.38***
	(0.013)	(0.07)	(0.019)	(0.10)
Observations	14,492	7,246	14,492	7,246

Notes:
Robust standard errors clustered at the community level in parentheses
*** p<0.01, ** p<0.05, * p<0.1
P-values are reported in parentheses for the Hansen J statistic and the C statistic
Ethiopia, Peru and India are country specific dummy variables
Country specific dummy variables suppressed
In column 2, IVs - two-period lagged HAZ, two-period lagged HAZ and its interaction with the country dummies two-period lagged hospital and its interaction with the country dummies, two-period lagged WAZ and its interaction with the country dummies

Table 4: Dynamic QR-IV Estimates

	HAZ	HAZ	HAZ	HAZ
	(1)	(2)	(3)	(4)
	QR-IV	QR-IV	QR-IV	QR-IV
	0.10	0.25	0.75	0.90
Lagged HAZ	0.82***	0.58***	0.14**	-0.15
	(0.11)	(0.08)	(0.06)	(0.15)
Ethiopia*lagged haz	-0.93***	-0.61***	-0.10	0.17
	(0.12)	(0.07)	(0.07)	(0.16)
India*lagged haz	-0.88***	-0.65***	-0.14	0.20
	(0.13)	(0.08)	(0.08)	(0.18)
Peru*lagged haz	-0.60***	-0.35***	0.12	0.46**
	(0.16)	(0.10)	(0.11)	(0.22)
Catch-up coefficient				
Ethiopia	-0.10***	-0.02	0.04*	0.016
	(0.02)	(0.02)	(0.02)	(0.04)
India	0.05	-0.07***	-0.003	0.05
	(0.04)	(0.02)	(0.04)	(0.05)
Peru	0.22**	0.22***	0.26***	0.31**
	(0.09)	(0.04)	(0.06)	(0.13)

Notes:
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
The full set of control variables as specified in column 2, Table 3 are included but suppressed
Ethiopia, Peru and India are country specific dummy variables

Table 5: Hypothesis testing: Dynamic QR-IV Estimates

Test	Q1=Q2	Q1=Q3	Q1=Q4
Ethiopia	-0.08***	-0.14***	-0.12**
	(0.02)	(0.03)	(0.05)
India	0.016	-0.05	-0.10*
	(0.03)	(0.05)	(0.06)
Peru	-0.006	-0.04	-0.08
	(0.08)	(0.10)	(0.16)
Vietnam	0.24**	0.68***	0.98***
	(0.09)	(0.13)	(0.18)

Notes:
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A1: Baseline differences in HAZ for attritors and non-attritors

Countries	HAZ Attritors (1)	HAZ Non-attritors (2)	Mean difference [1-2]
Ethiopia	-1.583	-1.581	-0.002
			(0.147)
N	197	1749	
India	-1.11	-1.35	0.24*
			(0.12)
N	170	1822	
Peru	-1.22	-1.30	0.08
			(0.10)
N	200	1840	
Vietnam	-1.52	-1.10	-0.42***
			(0.10)
N	155	1835	

Notes:

standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A2: Linear probability model of attrition

	Attrition Ethiopia (1)	Attrition India (2)	Attrition Peru (3)	Attrition Vietnam (4)
Height-for-age z score	-0.0008	0.009	0.012	-0.017
	(0.006)	(0.006)	(0.008)	(0.008)
Age in months	-0.003	0.005	-0.001	0.005
	(0.002)	(0.003)	(0.02)	(0.004)
Male dummy	-0.03	0.03	-0.07	0.035
	(0.05)	(0.05)	(0.05)	(0.06)
Community dummies	Yes	Yes	Yes	Yes
N	1946	1992	2040	1990

Notes:
standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1