Catch-up effects in health outcomes – Linear and Quantile Regression Estimates from four Countries*

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

September 2012
Work in Progress

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. The long-term detrimental consequences of such early childhood stunting is well established in the literature. 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 linear dynamic panel data models varies between 0 to 0.32, where, Ethiopia and India exhibit perfect catch-up and Peru and Vietnam exhibit partial catch-up in height-for-age z scores. To allow for the catch-up coefficient to vary along the entire distribution of child anthropometric outcomes, we also estimate a dynamic quantile regression instrument variable estimator. We find that the null of homogenous catch-up effects along the entire distribution of anthropometrics can be easily rejected.

*This study is based on research funded by the Bill & Melinda Gates Foundation (Global Health Grant OPP1032713) and Eunice Shriver Kennedy National Institute of Child Health and Development (Grant R01 HD070993). 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

¶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 individuals 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 growth in children can minimize the permanent effects of growth faltering. Existing estimates of 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 does early nutritional deficiency impact later health outcomes. In particular, what is the role of early nutritional deficiencies in explaining later life outcomes. The dynamic linear panel data regression models used in the literature capture average partial effects and do not allow for the impact of lagged health, socioeconomic characteristics, and individual characteristics to vary along the entire distribution of the child's current health status.

The objective of this paper is to fill this gap by estimating a dynamic quantile regression model, which allows the impact of the catch-up coefficient to vary along the entire distribution of anthropometric outcomes accounting for endogeneity in the lagged dependent variable.

Before identifying the varying distributional effects, we will first estimate a dynamic linear panel data model using first-difference generalized methods of moment (FD-GMM) estimator where the coefficient estimate on the lagged dependent variable captures the extent of recovery – complete, partial or none, from chronic malnourishment. First-differencing removes all time-invariant factors that are likely to be correlated with the lagged dependent variable in the right hand side (RHS) alleviating concerns of omitted variables bias and the instrument variables further addresses the problem associated with measurement error. The dynamic linear panel data model assumes that the catch-up coefficient has the same impact across the entire distribution of anthropometric outcomes. We test this assumption by estimating the Dynamic Quantile-Regression Instrument-Variable estimator, which allows the impact of lagged health (catch-up coefficient), household characteristics, individual characteristics and community characteristics to vary across the entire distribution of health outcomes. In particular, we are able to delineate the extent to which early nutritional deficiencies explain for poor health outcomes in the future.

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 - age 1, age 5, and age 8 in their lives. 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 40% of children measured at age 1 suffer from chronic nutritional deficiencies.

The FD-GMM estimator used here to compute the coefficient estimate on the lagged dependent variable in a linear dynamic panel data model results in a catch-up coefficient that varies from between 0 and 0.31. India and Ethiopia both exhibit almost 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 the 1% 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.

To allow for the potential of differential catch-up 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.

Our quantile regression instrument variable estimates suggest that children exhibit different levels of catch-up along the entire 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.

This paper contributes to the existing literature in many ways. First, the paper brings out the extent to which early nutritional deficiencies affect health status (height-for-age z-score) in later ages using a unique panel of children from four countries: Ethiopia, India, Peru, and Vietnam. This paper is the first to provide empirical evidence on catch-up effects in 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. In doing so, we treat the lagged dependent variable endogenous.

2 Data

The data used in this paper comes from the Young Lives Study, which measured heights and weights of children at age 1, that is, during 2002 in four different countries: Ethiopia, India, Peru, and Vietnam and then follows these children through second (2006-2007) and third (2009-2010) waves of the Young Lives Study. The panel nature of the data allows us to capture the transition in anthropometrics outcomes during three critical periods – at age 1, age 5, and age 8 in a child's life. The survey obtains detailed information on household demographic characteristics, income, assets, parental education, and height. The survey also includes a detailed community level questionnaire to capture access to paved roads, availability of health infrastructure, availability of drinking water, availability of electricity, distance to health center in km, and records prices on important basket of food items.[1] More details on the variables included in the empirical specification is provided in the next section.

The first wave of the Young Lives Study measured heights and weights of 1946 children in Ethiopia and of these, 1749 (10.12%) children are followed through the second and third waves of the Young Lives Study. In India, out of 1922 children surveyed during the first wave of the Young Lives Study, we are able to follow 1822 (8.53%) of these children through the second and third waves of the Young Lives Study. In Vietnam, out of 1990 children initially surveyed, we are able to follow 1835 (7.79%) children through the second and third waves of the Young Lives Study. In Peru, of the 2040 children initially surveyed in 2002, we are able to trace 1840 (9.80%) of these children through the second and third waves of the Young Lives 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 2006-07 and 2009-10.

A simple mean test on the difference in height-for-age-z score between attrited and non-attrited children is 0.002 (s.e. = 0.14) in Ethiopia, 0.24 (s.e. = 0.12) in India, 0.42 (s.e. = 0.10) in Vietnam, and -0.08 (s.e. = 0.10) in Peru. Baseline height-for-age z scores are not significantly different among attrited and non-attrited children in Ethiopia and Peru. However, there is some concern for the presence of attrition bias in Ethiopia and Vietnam. To further rule out attrition related selection bias, we follow, Mani, Hoddinott and Strauss (2012) and assume the following: (a) any information realized between periods t and t-2 that affects the probability of attrition is independent of the first-differenced error term in equation (2), (b) there is no serial correlation in the levels residuals specified in equation (1) – this allows two-period (and all further) lags of the health status to impact the attrition equation, and (c) the one-period lagged health has no impact on the attrition equation.

[1] We will be expanding on the use of community characteristics in the next version 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 ages of 1 and 8. Almost 30% of children at age 1 suffer from stunting and the worst incidence of stunting is observed in Ethiopia (46.31%). 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 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, incidence and intensity of chronic nutritional deficiencies has declined over time across countries as measured by both the percentage of children classified as stunted and in terms of mean height-for-age z-scores.

3 Empirical Specification

Fedorov and Sahn (2005)and Mani (2011) estimate the following dynamic linear panel data model where the coefficient 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].

$$H_{it} = \beta_0 + \beta_1 H_{it-1} + \sum_{j=1}^{R} \beta_j^X X_{jit} + \sum_{j=1}^{S} \beta_j^Z Z_{ji} + \epsilon_i + \epsilon_h + \epsilon_c + \epsilon_{it} \tag{1}$$

H_{it} and H_{it-1} are the child's HAZ scores measured at time t and t-1 respectively, where subscript i refers to the individual. X's are time-varying regressors which include child's age, household assets, and community characteristics such as availability of hospital, availability of health center, and availability of drinking water in the community. Z's include time-invariant characteristics such as gender.

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 β_1 is likely to be biased and inconsistent.

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

$$\delta H_{it} = \beta_1 \delta H_{it-1} + \sum_{j=1}^{R} \beta_j^X \delta X_{jit} + \delta \epsilon_{it} \tag{2}$$

OLS estimation applied to a FD 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 X's under the assumption of zero first-order and second-order serial correlation in the errors specified in the levels specification (1).

Standard linear specifications assume that the catch-up coefficient has the same impact across the entire distribution of anthropometric measurements. To allow and test for this possibility we will also estimate the following dynamic panel data model using a Quantile-Regression framework as described below:

$$Q(\tau)_{hit} = \beta_0 + \beta_1(\tau)H_{it-1} + \sum_{j=1}^{R} \beta_j^X(\tau)X_{jit} + \sum_{j=1}^{S} \beta_j^Z(\tau)Z_{ji} + \epsilon(\tau)_{it} \tag{3}$$

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 the different quantiles. To address the endogeneity in the lagged dependent variable in equation (3), we will estimate the model using a Quantile-Regression-Instrument-Variable (QR-IV) estimator proposed by Galvao (2011).

The outcome variable of interest in this paper is height-for-age z-score (HAZ), a well established long-run indicator of individual health status. 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 drinking water, availability of health center, and availability of hospital. The control variables are chosen to reflect standard models of determinants of human capital accumulation in child health [Thomas and Strauss, 1992, 2008]. Descriptive statistics on the outcome variable and all the regressors used in the empirical specification are summarized in table 2.

4 Results

The OLS results from estimating the dynamic linear panel data specified in equation (1) is reported in column 1, table 3. OLS estimate on the one-period lagged height-for-age z score is 0.66 for Vietnam (see column 1, table 3), this indicates less than partial catch-up in attained height-for-age z scores. We observe higher levels of catch-up potential in Ethiopia, India, and Peru. The OLS estimate is 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 dummy capturing the incidence of diarrhea during the last 24 hours, the interaction of this variable with the country specific dummies, two-period lagged dummy capturing the incidence of diarrhea during the last 24 hours interacted with father's education, and the interaction of this variable with the 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 0.23 in Vietnam, 0.005 in Ethiopia, -0.03 in India, and 0.31 in Peru. We find that the null of perfect catch-up cannot be rejected at the 1% significance level for children in Ethiopia and India. We find evidence in support of partial catch-up among children in Vietnam and Peru, again these estimates are statistically significant at the 1% significance level.

The coefficient on the catch-up term from the first-difference GMM 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.

To allow for the potential of differential catch-up 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 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.

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 dependendence 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

It is well known that if the correlation between the endogenous regressor and the instruments is weak, IV estimates remain inconsistent [Murray (2006)]. We use the Kleibergen-

Paap Wald rk F statistic to test for weak instruments. This test statistic is robust in the presence of heteroskedasticity, autocorrelation and clustering [Kleibergen and Paap, 2006]. The Kleibergen-Paap Wald rk F statistic on the excluded instruments reported in our Arellano-Bond estimates is always above 10, satisfying the Staiger and Stock (2003) rule of thumb rejecting the null of weak correlation between the instruments and the endogenous regressor.

Arellano and Bond (1991) stress that in their estimator, using a twice lagged dependent 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 column 3, table 3. At the 5% 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.

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) of 2.33 with a p-value of 0.67 (column 2, table 3) suggests that we cannot reject the null of instrument validity for the instruments specified in columns 2, table 3. The coefficient estimate on the Hansen J statistic and the first-stage F test statistic on the excluded instruments are all appended at the end of the regression table 3.

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

6 Conclusion

In this paper we estimate - (a) the 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 - age 1, age 5, and age 8 in their lives. 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. Stunting is a serious source of concern among policy makers in these countries, more than 40% of children measured at age 1 suffer from chronic nutritional deficiencies.

The FD-GMM estimator used to compute the coefficient estimate on the lagged dependent variable in a linear dynamic panel data model results in a catch-up coefficient that varies from between 0 and 0.31. India and Ethiopia both exhibit almost 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 the 1% 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.

To allow for the potential of differential catch-up 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.

This paper contributes to the existing literature in many ways. First, the paper brings out the extent to which early nutritional deficiencies affect health status (height-for-age z-score) in later ages using a unique panel of children from four different countries Ethiopia, India, Peru, and Vietnam. This paper is the first to provide empirical evidence on catch-up effects in 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.

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 on Height-for-age z-score

Panel A: Pooled sample - Ethiopia, India, Vietnam, and Peru			
Years	% HAZ <-2	Mean	Mean difference
2002	31.11	-1.40	-0.09*** (2006-2002)
		(0.01)	(0.02)
2006	30.91	-1.49	0.26*** (2009-2006)
		(0.01)	(0.17)
2009	22.35	-1.23	0.17*** (2009-2002)
		(0.01)	(0.02)
Panel B: Ethiopia			
Years	% HAZ <-2	Mean	Mean difference
2002	46.31	-1.81	0.33*** (2006-2002)
		(0.03)	(0.04)
2006	31.16	-1.48	0.25*** (2009-2006)
		(0.02)	(0.03)
2009	20.90	-1.22	0.60*** (2009-2002)
		(0.02)	(0.04)
Panel C: India			
Years	% HAZ <-2	Mean	Mean difference
2002	30.13	-1.35	-0.30*** (2006-2002)
		(0.03)	(0.04)
2006	35.40	-1.65	0.19*** (2009-2006)
		(0.02)	(0.03)
2009	29.08	-1.45	-0.10*** (2009-2002)
		(0.02)	(0.04)
Panel D: Vietnam			
Years	% HAZ <-2	Mean	Mean difference
2002	21.35	-1.13	-0.20*** (2006-2002)
		(0.02)	(0.03)
2006	24.45	-1.33	0.23*** (2009-2006)
		(0.02)	(0.03)
2009	19.34	-1.10	0.04 (2009-2002)
		(0.02)	(0.03)
Panel E: Peru			
Years	% HAZ <-2	Mean	Mean difference
2002	27.52	-1.32	-0.20*** (2006-2002)
		(0.02)	(0.03)
2006	32.70	-1.53	0.38*** (2009-2006)
		(0.02)	(0.03)
2009	20.65	-1.15	0.18*** (2009-2002)
		(0.02)	(0.03)

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

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

Table 2: Summary statistics of all variables used in the empirical specification

Variables	Observations	Mean	Std. dev
Height-for-age z-score (HAZ)	21738	-1.37	1.17
Age in months	21738	57.45	34.89
Illness during the last 24 months in wave 1	7246	0.14	0.34
Diarrhea during the last 24 months in wave 1	7246	0.04	0.20
Asset index	21738	0.001	2.08
Male dummy	21738	0.52	0.49
Rural dummy	21738	0.62	0.48
Dummy for the availability of electricity	21738	0.76	0.42
Dummy for the availability of hospital	21738	0.47	0.49
Dummy for the availability of health center	21738	0.86	0.34
Dummy for the availability of drinking water	21738	0.58	0.49

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

Table 3: Linear Dynamic Panel Data Estimates of the Catch-up Coefficient

Variable	HAZ (1) OLS	HAZ (2) FD-GMM
Lagged HAZ	0.664***	0.238***
	(0.0126)	(0.08)
Ethiopia*lagged haz	-0.229***	-0.232***
	(0.0187)	(0.08)
India*lagged haz	-0.0997***	-0.275***
	(0.0180)	(0.09)
Peru*lagged haz	-0.0780***	0.08
	(0.0166)	(0.11)
Male dummy	0.112***	
	(0.0238)	
Lagged age in months	0.00789***	-0.008**
	(0.000605)	(0.004)
Lagaged age in months*male dummy	-0.00281***	-0.0012***
	(0.000488)	(0.00032)
Rural dummy	-0.174***	-0.21***
	(0.0283)	(0.03)
Asset index	0.0608***	0.016**
	(0.00579)	(0.0074)
Dummy for the availability of electricity	-0.0485	0.22
	(0.0446)	(0.17)
Dummy for the availability of hospital	0.109***	0.0538
	(0.0323)	(0.0783)
Dummy for the availability of health center	0.0887***	-0.003
	(0.0336)	(0.12)
Dummy for the availability of drinking water	-0.0859***	-0.152*
	(0.0288)	(0.08)
Kleibergen-Paap F statistic		13.79
Hansen J statistic		11.05
		(0.20)
Test of linear hypotheses		
Lagged HAZ+Ethiopia*lagged haz		0.005
		(0.02)
Lagged HAZ+India*lagged haz		-0.03
		(0.03)
Lagged HAZ+Peru*lagged haz		0.31***
		(0.07)
Observations	14,492	7,246

Robust standard errors in parentheses

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

P-values are reported in parethesis for the Hansen J statistic

Ethiopia, Peru and India are country specific dummy variables

Table 4: Dynamic Quantile Regression Instrument Variable Estimates of the Catch-up coefficient

Variable	HAZ (1) QR-IV 0.10	HAZ (2) QR-IV 0.25	HAZ (3) QR-IV 0.75	HAZ (4) QR-IV 0.90
Lagged HAZ	0.90***	0.58***	0.15**	-0.20
	(0.013)	(0.08)	(0.07)	(0.15)
Ethiopia*lagged haz	-1.03***	-0.58***	-0.07	0.42**
	(0.015)	(0.08)	(0.08)	(0.17)
India*lagged haz	-0.97***	-0.57***	-0.20**	0.15
	(0.15)	(0.09)	(0.09)	(0.18)
Peru*lagged haz	-0.74***	-0.35***	0.07	0.58**
	(0.18)	(0.11)	(0.11)	(0.24)

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Ethiopia, Peru and India are country specific dummy variables