

The Effect of Child Health on Schooling: Evidence from Rural Vietnam

by

Thuan Quang Thai *

and

Evangelos M. Falaris**

January 2011

*Thai: Max Planck Institute for Demographic Research, Konrad-Zuse-Strasse 1, 18057 Rostock, Germany; e-mail: thai@demogr.mpg.de; telephone: + 49 (0)381 2081-236.
Falaris: Department of Economics, University of Delaware, Newark, DE, 19716, USA; e-mail: falaris@udel.edu; telephone: +1-302-831-1768. We thank Simon Condcliffe, Saul Hoffman, James Mulligan, and Peter Schnabl for their comments and the General Statistical Office of Vietnam for making the data available.

Abstract

We study the relationship between long term child health and human capital. Child health may suffer if a child is inadequately nourished or is exposed to disease early in life and this may affect subsequent accumulation of human capital. We use data from rural Vietnam to examine the impact of child health on delay in starting school and schooling progress taking into account that choices of families affect children's health and schooling. Our instrument is early life rainfall shocks that have differential effects arising from regional economic diversity. Our estimates indicate that better child health results in meaningfully improved schooling outcomes.

Keywords: child health, z-score, school entry delay, schooling gap, rainfall shocks, Vietnam.

JEL classifications: I12, J24, J13, O15.

Introduction

Long term child health, which is important for the full development of a child's capacities, is likely to suffer if a child is not adequately nourished or is exposed to disease¹ during the early years of life. The aspects of human capital investment that we study are a child's delay in starting school and the child's progress through school. The relation between long term child health and schooling is of interest in the assessment of a policy that results in an improvement in child nutrition which improves child health. Better child health is a worthwhile goal in itself. Better child health may also facilitate human capital investment which will promote economic growth. Our study has implications for evaluating such longer term economic effects.

We measure long term child health by standardized height-for-age and specify and estimate empirical models of the relationship of height-for-age and measures of school entry delay or progress through school. We take into account that choices made by families affect both their children's health and schooling. We estimate our models with data from rural Vietnam where shortfalls in long term child health are a significant problem whose amelioration is a vital policy concern. We estimate our models by simultaneous equation methods, using early life rainfall shock as an instrument. We estimate separate rainfall shock effects by geographic region. Our estimates of the effects of rainfall shocks take into account and reflect the regional diversity of economic activity in rural Vietnam.

¹ For recent findings on the effect of exposure to disease in early life see Strauss and Thomas (2008). Glewwe and Miguel (2008) discuss the evidence on the effect of early life nutrition and exposure to disease on long term child health.

We find that better long term child health results in better schooling outcomes. The simultaneous equation estimates are the preferred specification and are much bigger than estimates of a model that ignores the endogeneity of long term child health. Our estimates are both statistically and empirically important and indicate that better long term child health results in shorter school entry delays and more rapid progress through the grades. They imply that a realistic policy intervention that improves long term child health would result in improved human capital investment outcomes and have a substantial effect on wages. Notable among our results is that we do not find significant differences by gender in the effect of child health in the outcome equations. This finding is in contrast with significant gender differences obtained in previous studies for several other developing countries. We also find that members of ethnic minority groups suffer delayed school entry and slower progress through the grades.

Framework

Existing models of the relationship between health and human capital investment such as Glewwe and Miguel (2008) predict that parents with characteristics that result in higher investment in their children's long term health (henceforth child health) will also invest more in the children's human capital. According to this model, a healthier child develops better cognitive ability and is more ready to start school at the conventional age and make academic progress relative to a less healthy child.

Other models, such as Rosenzweig (1990), consider child labor and investment in children's human capital as competing activities. Thus, in this model, an increase in child wages (the opportunity cost of time spent investing in human capital) is predicted to result in a reduction in the time spent investing in human capital. This model does not

treat investment in child health as endogenous, but its argument can be extended to the more general case of endogenous child health: families with characteristics such that their child enjoys better health may find that better child health results in higher child labor productivity. This higher productivity increases the opportunity cost of investing in human capital resulting in reduced human capital investment. This last effect is likely to be particularly strong among credit-constrained families such as small farmers. In this setup, better child health increases the opportunity cost of being in school and is expected to result in reduced child human capital investment.

Considered together, these models imply opposing effects of better child health on human capital investment: readiness/progress in school and opportunity cost through the reward for child labor. Which effect will be predominant is an empirical matter. The gist of these arguments is that child health should be treated as a choice variable in regressions that have child schooling outcomes as the endogenous variable. The sign and extent of the bias from ignoring the endogeneity of child health is an empirical matter.

Measures of health such as child height may be subject to measurement error. Child height is a proxy for child health, which it measures with error. Child health itself may be measured with error. The survey from which our data are drawn has been conducted with reduction of measurement error as an important goal (Grosh and Glewwe, 2000), World Bank (2001) but some measurement error may be present anyway. Measurement error in a regressor can be expected to lead to attenuation bias of its OLS coefficient. To the extent that the survey organization's efforts to reduce measurement error have been successful, we expect the degree of attenuation bias to be moderate.

In summary, we expect that an OLS regression of a measure of child schooling on a measure of child health will be biased (up or downward) as a result of behavioral responses of families and will be biased downward (perhaps moderately) because of measurement error. A way to overcome these biases is to use a simultaneous equation method as we discuss in the next section.

A number of previous studies of the effect of child health on schooling outcomes have generally found that the effect of child health is biased toward zero if the endogeneity of child health is ignored. These studies include Glewwe and Jacoby (1995), Alderman et al. (2001), (2009), Glewwe, Jacoby, and King (2001), Hoddinott and Kinsey (2006), and Yamauchi (2008). Some of these studies rely on longitudinal data which is not a practical option for us. The instruments used in some previous studies such as price shocks or mother's height may be of questionable validity as argued by Glewwe and Miguel (2008). Maccini and Yang (2009) and Hoddinott and Kinsey (2001) established that early life exogenous shocks such as rainfall shocks have effects on health and on adult schooling and economic outcomes. We also use rainfall shocks as an exogenous instrument and adapt this measure to the characteristics of the agricultural economy of Vietnam.

Econometric Model and Data

Econometric Model

The empirical model we estimate is as follows:

$$Y_{ijkt} = \alpha_1 h_{ijkt} + \beta_1 Z_{ijkt} + \gamma_1 X_{jkt} + f_{1kt} + v_{1ijkt} \quad (1)$$

$$h_{ijkt} = \beta_2 Z_{ijkt} + \gamma_2 X_{jkt} + \delta_k S_{ijkt} + f_{2kt} + v_{2ijkt} \quad (2)$$

Y_{ijkt} is a measure of an educational outcome of an individual child i born in year t , who lives in community j , region k . h_{ijkt} is health stock measured at the minimum enrollment age. We are interested in estimating the treatment effect of child health represented by the coefficient α . Z_{ijkt} is a vector of observed individual child characteristics such as age, gender, and ethnicity, family characteristics including parents' education, number of siblings, and family income. X_{jkt} is a vector of community characteristics such as schooling infrastructure. v_{1ijkt} is the error term representing those unobserved variables and heterogeneities such as parental tastes for child health and education, child specific attributes, as well as measurement error.

The regression model also includes the fixed effect f_{kt} . This term is intended to capture unobserved factors that vary by region and birth cohort and that are possibly correlated with the other regressors. The regional fixed effects do not control for family unobservable factors. To deal with the possible presence of such factors we specify equation (2). In equation (2) child health (h_{ijkt}) is a function of the same regressors as appear in equation (1), exogenous instruments S_{ijkt} which we specify to measure rainfall shocks with separate coefficients by region, and an error term v_{2ijkt} .

If we were to ignore the possible correlation of the child health measure with the error term, we would estimate equation (1) by ordinary least squares and allow for fixed effects (OLS-FE). However, as discussed in the previous section, h_{ijkt} may be correlated with v_{1ijkt} . More precisely, h_{ijkt} is correlated with v_{2ijkt} . If v_{2ijk} is correlated with v_{1ijk} because of unobserved omitted variables that affect both, then single equation estimates will suffer from simultaneous equation bias.

The correlation of the error terms may occur for several reasons. Parents who value their children's health may also value their education. This would result in omitted variables bias in equation (1). Another possible pattern of parental behavior that would bias the single equation estimates of equation (1) is if parents smooth health outcomes among their children. If a child's baseline healthiness is observed by parents but not the researcher, and if parents direct more resources towards their less healthy children, the OLS-FE estimates will be biased. As discussed in the previous section, labor that could be performed by children (e.g. as helpers in the household or family farm) could play a role in children's educational outcomes and this leads to a further potential of omitted variable bias: an increase in characteristics of a child that result in better health, will also increase the opportunity cost of attending school. For a credit and resource-constrained family, there would be incentives to postpone or reduce investments in a child's schooling if that child could contribute relatively more to family income in the short run. Lastly, despite the survey organization's efforts to reduce survey error, there could be measurement error in child height, the reporting or recording of a child's birth date or other regressors. Measurement error results in attenuation bias of the OLS-FE estimates. The direction of the bias in single equation estimates that would arise from the families' behavioral responses can be upward or downward. Measurement error results in bias toward zero. Estimating equations (1) or (2) by a simultaneous equation method such as two stage least squares (2SLS-FE) enables us to overcome and measure these possible biases that may arise if we ignore the endogeneity of child health.

We use the variation in rainfall during a child's birth year as our instrument and we adapt it to the context of Vietnam by allowing for different effects of variation in

rainfall by region. We hypothesize that there are two main channels through which the variation in rainfall could be linked to child health. First, rainfall is likely to affect child health through its effect on family income. The livelihood of many rural Vietnamese households depends on short term crops such as rice. Stable rainfall is one the most important determinants of a successful rice harvest. Therefore, shocks in rainfall such as floods or droughts are adversely related to household income and nutritional conditions of mother and infants. Second, shocks in rainfall are also associated with deterioration in the health environment, which affects child health. Many communicable diseases that affect infants such as respiratory infections, Japanese encephalitis, and others are usually more prevalent following extreme drought or flooding (World Health Organization, 1999). With respect to the income channel mentioned above, it is necessary to assume that the household has no ability to access temporary credit or transfers in coping with an undesirable shock in rainfall. We elaborate on this point as it applies to rural Vietnam in the next section.

In addition to a child's health, household income might be another determinant of schooling outcomes. A low income family faced with educational costs such as school fees, clothes, and books may choose to delay or curtail a child's schooling in view of the fact that the child may be able to assist in some of the family's productive activities such as babysitting or field work. Moreover, if a previous health shock has a long term effect on present income, the inclusion of income in the outcome equation controls for this correlation. However, inclusion of family income in the schooling outcome equations may create another source of bias because of simultaneity. It is possible that parents give priority to income-generating activities and delay school entry

or restrict the schooling attainment of their children. Schooling outcomes in this case might have a causal effect on family income rather than the other way around. We explore the effect of the possible endogeneity of income in the empirical analysis.

Data

We use the 1998 Vietnam Living Standard Survey (VLSS) data in the analysis. In order to obtain estimates that are comparable to those in previous studies, we consider only children currently living in rural areas with both parents. The VLSS survey first selects communes, which is the lowest administrative level in the Vietnamese administrative system. Observations (households) are, then, drawn from two randomly selected villages in each commune. Since Vietnamese households tend to share similar socioeconomic conditions at the commune level, it is reasonable for the standard errors to be clustered at the commune level in the estimation. The sample used for the estimation covers 130 rural communes located in 7 different regions throughout the country. In general, the definition of regions reflects differences in climatic, topographic, historical and socioeconomic characteristics.

We examine the impact of child health on two schooling outcomes: age at entry to school and years of completed schooling. In the Vietnamese education system, the minimum enrollment age is six. The age at which a child is first enrolled was explicitly asked during the interview so the information on enrollment status is directly obtained. Therefore, our first schooling measure, years of school entry delay, is calculated by subtracting six from the observed age at entry.² The second schooling

²It is possible that a child older than six years old is still not enrolled as of the interview date and information on delay is right censored by age minus six. Having conducted a check on the selected sample,

outcome that we wish to study is years of schooling attainment. This variable is likely to be censored. If a child is still in school when the survey ends, information on the child's highest grade is censored by the current grade. To overcome this problem, we measure schooling attainment by grade progress. Those children who experienced adverse earlier childhood health conditions not only experience a delay in starting school, but also struggle with grade completion. Falling behind the normal progress results in schooling gaps and the analysis attempts to relate child health to this gap. The schooling gap is defined as $Age - current\ grade - five$. Since we can observe the current grade for all children who are still at school, the schooling gap is not subject to censoring. For school leavers, the schooling gap is measured as $Age - highest\ grade - six$. Highest grade is calculated as follows. If a child is not currently attending school, then information on the last schooling level is obtained. The highest grade of this schooling level is assigned if the child either obtained a diploma or completed the last grade for the respective level. Otherwise, the highest grade completed before leaving school is assigned.

Variables that measure individual child characteristics in the VLSS include age, ethnicity, gender, and health status. The health economics literature argues that health status is multidimensional and different health indicators have different interpretations (Strauss and Thomas 1998). It is recognized that health measures such as disease incidence, Body Mass Index (BMI) or child weight-for-age are likely to affect schooling outcomes. However, these indicators reflect short run health status and are time-varying. These time-varying health measures make it difficult to pin down the health effect. Height attained, on the other hand, reflects the long run dimension of health

we found that those children who report that they were not enrolled are those school drop-outs. This fact allows us to infer that years of entry delay is not likely to be right censored.

and is appropriate for an analysis whose focus is the investment decisions of parents in their children's human capital.

We follow a widely used approach in the literature and use height-for-age as an indicator of child health.³ An anthropometric expert in the VLSS survey team measured directly the height of respondents. The VLSS measures the ages of respondents in years for those greater than or equal to age ten and in both years and months for those less than age ten. We round the age in months of those individuals less than age ten to their age in completed years.⁴ For instance, 7 years and 10 months is treated as age 7. Admittedly, a certain degree of error arises from this rounding for part of the sample with ages between 6 and 9 years old. However, the sample includes less than 25 percent of these individuals.⁵

Family characteristics include parents' schooling, family income per capita and number of siblings. Family income per capita is represented by the logarithm of food expenses per capita during the last 12 months, adjusted for inflation and using a location-specific price index. Community characteristics include three schooling quality variables: number of teachers per class, percentage of classrooms in poor condition and distance to nearest primary school.

³ We transform height-for-age into z-score using the reference height-for-age distribution of children in the United States which is published by the National Center for Health Statistics. A child is considered stunted if his or her z-score is below -2. Assuming that there is no considerable change in the child health stock after age 6, we use z-scores at ages older than six to approximate the health stock at age 6 in our empirical work. This approximation is also based on the fact that the z-score is a standardized indicator of health that is comparable across ages and genders.

⁴ Measuring the z-score based on age in months does not significantly affect our estimates. See also footnote 13.

⁵ There are 1100 children ages 6-9 whose age in months is greater than 6.

Rainfall

We use information on historical shocks in rainfall in Vietnam in order to construct an exogenous instrument for child health. We have no access to data on historical precipitation recorded at local observatory stations. Instead, we use estimated precipitation developed by climate researchers (Legates and Willmott, 1990).⁶ Rainfall data come from the Gridded Monthly Time Series (Version 2.01) dataset, henceforth, GMTS. This dataset contains global historical estimates of rainfall for a grid of 0.5 degree by 0.5 degree of latitude/longitude, where the grid nodes are centered on 0.25 degree. Thus, the area covered by each grid is approximately 50 square kilometers. This is also the maximum distance between a commune and its closest grid point of rainfall. To derive the variable rainfall shock, we matched each community in the VLSS sample with the four closest grid points in the GMTS dataset. Since there was no information on longitude and latitude for each community in the version of the VLSS available to us, we determined its geographical location using the administrative map provided by the Government of Vietnam.⁷ Rainfall is averaged over the four grid points to be used as commune-specific rainfall data.

In calculating the child's birth year rainfall we follow the method described in Maccini and Yang (2009), using the total twelve months of rainfall (in mm) for the two consecutive wet and dry seasons, starting with the season in which a child is born. Some regions in Vietnam have all four seasons but this seasonality is mostly related to temperature and number of sunny days. In terms of precipitation, all regions experience

⁶ The dataset is provided by Center for Climatic Research, Department of Geography, University of Delaware. Terrestrial Precipitation: 1900-2008 Gridded Monthly Time Series - Version 2.01, interpolated and documented by Kenji Matsuura and Cort J. Willmott (with support from IGES and NASA).

⁷ For further details, visit <http://gis.chinhphu.vn/>. The map for sample communities and closest grid points is available from the authors upon request.

wet and dry seasons with sharp differences in monthly rainfall. The wet season includes months with a minimum rainfall of 100 mm. We use the information on the month of birth reported in the VLSS survey in comparison with the local seasonality to determine the birth season for an individual child. Mean annual rainfall for each commune is calculated over a 49-year period from 1950 to 1999. Finally, the birth year rainfall shock is defined as the percentage change of birth year rainfall over the mean annual rainfall in the child's community.⁸

Definition of Instrumental Variable

The shock in rainfall is likely to influence child health via the income and health environment channels as mentioned earlier. This impact is not likely uniform throughout the country, but varies among regions. The definition of regions in Vietnam takes into account geographical features. For instance, some regions face the sea while other regions are inland. Some regions are mountainous while other regions are mainly lowland. These geographic features result in different climatic characteristics and thus a different pattern of annual rainfall by region. For instance, the regions of Red River delta and Northern Central are influenced by a monsoon wind pattern that tends to make rainfall in wet and dry seasons more extreme. The region of the Mekong delta has different ecological characteristics from the other regions. Rather than depending on rain, the Mekong delta is a flat estuary area of the Mekong River. This region experiences seasonal flooding when the upstream areas have their wet seasons.

The second reason for possible heterogeneous impact of rainfall by region has to do with the regional agricultural production pattern. Crop production is not the

⁸ This rainfall shock is therefore equal to the difference in the logarithms of birth year rainfall and mean annual rainfall.

same in all regions of Vietnam. It is likely that in places where production of short term rain-fed crops takes place, the impact of a shock in rainfall would be the most pronounced. We expect that the impact of rainfall on life outcomes in these regions would be different from that in the remaining regions. Therefore, we interact the variable measuring birth year rainfall shocks with regional dummies. These interactive rainfall shock variables are our instrumental variables.

Farmers

During the relevant period in our sample (1978-1992) small, poorer farmers in Vietnam, many of whom grew short term rain-fed crops, had limited access to credit that they could use to smooth consumption in the face of an adverse rainfall shock. In the period prior to 1988 the majority of farmland in Vietnam had been collectivized and was cultivated inefficiently. However, many farmers also cultivated small private plots but had no access to credit. In 1988 collectivization was reversed and farmers were given long term use rights over land that was allocated to them. Land could not be officially mortgaged. The Land Law of 1993 (after the birth of the youngest children in our sample) permitted the mortgaging of land use rights. As part of the banking reforms of the late 1980s, the Agricultural Bank of Vietnam was established. However it dealt mainly with larger commercial farm operations and did not provide much credit to small, poorer farmers. A subsidized interest rate arm of the Agricultural Bank of Vietnam, the Bank for the Poor, was established in 1995 (after the relevant period in our sample) in order to serve small, poorer farmers (Wiens, 1998; Ravallion and Van de Walle, 2008). As of 1993, there was evidence that poorer Vietnamese farmers had a lower value of

borrowing (and even lower value of liquid assets) relative to land than richer farmers (Wiens, 1998).⁹

Small, poorer farmers could borrow from private moneylenders at a high cost. Borrowing from relatives many of whom may have very low incomes was often not feasible. Therefore, a farmer who wished to obtain a loan in order to smooth consumption had to arrange it with a supplier or buyer of his output.

Suppliers or buyers would be more likely to extend credit to a farmer who was likely to have a higher ability to repay the loan. Larger cash-crop farming operations in many regions of Vietnam and larger rice farms in the Mekong Delta were more likely to have the future ability to repay. In addition such larger farms typically had a continuing business relationship with suppliers or buyers which reinforced the motivation to repay. Small rice-growing farms, on the other hand (mainly in the Red River delta and North Central regions) sold output in farmers' markets to small-scale traders who had relatively lower ability to extend credit to them and the small farmers themselves could reasonably be considered as having a relatively lower future ability to repay a loan. Thus, operators of larger farms could more easily smooth consumption than operators of small farms¹⁰.

An additional way in which a rural family could mitigate the effect of an adverse rainfall shock may be by moving to a more favorable location. However, Vietnam had strict controls on geographic mobility until the early 1990s (Wiens, 1998). Thus a rural family could not readily respond to an adverse rainfall shock by moving.

⁹ In 1998 there were some poverty alleviation programs in rural Vietnam. Their focus was on microfinance, extension services, and infrastructure development rather than relief in the face of natural disasters.

¹⁰ Livestock operations that farmers would have in addition to their farm were more likely to be commercial operations for larger farms that could help smooth consumption if farm output suffered an adverse shock. In contrast, small farmers would raise some livestock and use farm byproducts as animal feed and had limited ability to use market inputs. If a small farm's output suffered an adverse shock, the livestock operation suffered a similar shock and farmer income suffered accordingly.

Descriptive Statistics

Summary statistics for the VLSS sample are presented in Table 1. After

Table 1: Summary statistics, communities, families and children ages 6 – 19 in rural Vietnam, VLSS 1998

Variable	N	Mean	Std. Dev.	Min	Max
Delay (1 if yes)	4754	0.28	0.45	0.00	1.00
Years of entry delay	4754	0.41	0.80	0.00	7.00
Years of schooling gap	4754	1.55	1.92	0.00	12.00
Highest grades	878	6.19	2.51	1.00	12.00
Repeat primary grades	4754	0.19	0.39	0.00	1.00
Repeat lower secondary grades	4754	0.02	0.15	0.00	1.00
Height-for-age (z-score)	4754	-1.70	0.96	-4.88	3.26
Stunted	4754	0.38	0.48	0.00	1.00
Gender (1 if male)	4754	0.52	0.50	0.00	1.00
Age	4754	12.03	3.63	6.00	19.00
Ethnicity (1 if Kinh)	4754	0.88	0.33	0.00	1.00
Mother's schooling	4754	6.50	3.00	1.00	15.00
Father's schooling	4754	7.55	2.97	1.00	19.00
Number of siblings	4754	2.99	1.24	1.00	7.00
Log of per capita expenditure	4754	7.04	0.36	5.64	9.59
Distance to nearest primary school (km)	4754	0.70	1.03	0.00	5.00
Teachers per class	4754	1.27	0.20	0.88	1.95
Fraction of classrooms in poor cond.	4754	0.36	0.22	0.00	1.00
Northern mountain	4754	0.19	0.39	0.00	1.00
Red river delta	4754	0.19	0.39	0.00	1.00
North Central	4754	0.16	0.37	0.00	1.00
Southern Central	4754	0.10	0.30	0.00	1.00
Central Highlands	4754	0.08	0.27	0.00	1.00
Southeast	4754	0.10	0.30	0.00	1.00
Mekong delta	4754	0.18	0.38	0.00	1.00
Birth year rainfall shock (% from mean)	4754	-0.24	0.33	-1.90	0.35

eliminating all observations with missing values, the sample that is used in the estimation

includes 4754 children with ages from 6 to 19 years. The proportions for boy and Kinh

majority ethnicity are 52% and 88% respectively. We eliminated extreme outliers of

height-for-age z-scores so that the range of this variable in the sample is restricted from -5 to 5.¹¹ On average, the standardized height-for-age for Vietnamese children is about 1.7 standard deviation below the median of the reference height-for-age distribution. In the sample 38% of children are categorized as stunted compared to the height-for-age distribution of the United States. This stunting rate is not very different from the rate of 35% calculated for children ages 0 to 60 months with the same data by Glewwe (2004).

The last line of Table 1 shows summary statistics of the rainfall shock variable. In general, it seems that the average child in the sample experienced drought conditions. On average, the birth year rainfall shock is about 24% below the annual mean rainfall in the birth district. The median child experiences a rainfall shock in the birth year of 19% below the annual mean. This particularly low rainfall can be linked to the fact that Vietnam is one of the countries influenced by the recurrent impact of the El Niño Southern Oscillation (ENSO) climate phenomenon. An ENSO happens every 2 to 7 years and is typically associated with anomalies in precipitation resulting in severe drought in the case of Vietnam (United Nations, 2000). In our sample, which includes individuals born between 1978 and 1992, three strong episodes of ENSO have been recorded in the years 1982-83, 1986-87 and 1991-92 (Trenberth, 1997). The low rainfall level during these episodes was not offset by sufficient high rainfall episodes in the sample period. In view of these facts, we consider the negative values of the sample mean and median birth year rainfall shock¹² to be plausible.

¹¹ We only lose four observations of the edited sample because of this restriction.

¹² It should be noted that the Indonesian sample used in Maccini and Yang (2009) covers a longer range of birth cohorts (1953-1974) and without unusually pronounced drought episodes. Their sample average rainfall shock is therefore closer to zero.

Estimates

Effect of Rainfall on Child Health in Rural Vietnam

In order to analyze the effect of rainfall on child health, we present separately the first stage regression in Table 2, in which the height-for-age z-score is regressed on the selected instruments and other regressors. As explained earlier, the first stage regression also controls for fixed effects of birth year interacted with geographic regions. In preliminary regressions that controlled for the effect of rainfall shocks in the seven geographic regions into which Vietnam is divided, it became apparent that some of the region-specific effects of rainfall shocks were very similar among subsets of regions. Therefore, we carried out a statistical test of whether or not it is appropriate to restrict the regional rainfall shock effects to be equal for some regions.¹³ Based on the outcome of the test we estimate five regional rainfall shock effects.

To examine the link between family income, child health and child schooling progress, the first regression in Table 2 includes the variable of per capita food expenditure in logarithmic form. Food expenditure is likely to be a more accurate measure of the resources available to a family than reported income. Because no information on family food expenditure is available at the time a child was very young (e.g. two years old or younger) and at the time parents make the educational decision for their child, we use current (1998) food expenditure as a proxy for past income. The proxy

¹³ We conducted an F- test of the equality of the rainfall shock effects among the Red River and North Central regions and also among the South Central and Central Highlands regions. We have $F(2, 129) = 0.23$, $Pr = 0.798$ for the model without income. According to the test we cannot reject the null hypothesis of the equality of each pair of coefficients and thus we estimate common coefficients for each pair of regions.

would be a good one provided that rural households experienced relatively little growth in income in the 1980s and 1990s despite the economic growth that occurred in Vietnam during that period.

Table 2 – Estimates of the Child Height for Age Equation
Dependent Variable: Height-for-Age z-score

Variables ^a	With p.c. food expenditure		Without p.c. food expenditure	
	b	se ^b	b	se ^b
Gender (1 if male)	-0.174***	0.03	-0.168***	0.03
Ethnicity (1 if Kinh)	-0.039	0.07	-0.038	0.08
Father's schooling	0.01	0.01	0.016*	0.01
Mother's schooling	-0.007	0.01	-0.003	0.01
Number of siblings	-0.027	0.02	-0.046**	0.02
Distance to nearest primary school	-0.033	0.02	-0.034	0.02
Teachers per class	0.221	0.16	0.245	0.17
Fraction of poor classrooms (%)	0.062	0.11	0.082	0.11
Log of per capita food expenditure	0.343***	0.07		
Rain shock in northern mountain	-0.018	0.17	-0.049	0.17
Rain shock in Red River & North Central	0.443*	0.19	0.405*	0.19
Rain shock in South Central, Highlands	-0.500*	0.22	-0.516*	0.23
Rain shock in Southeast	-1.164***	0.16	-1.115***	0.17
Rain shock in Mekong Delta	-0.423**	0.16	-0.395*	0.15
F(5, 129) ^c	13.92		11.28	
p-value	0		0	
N	4754		4754	

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

^a An overall intercept and fixed effects for region and birth year are also included.

^b The estimates of the standard errors take account of the clustering of observations by commune.

^c F test of the hypothesis that the coefficients of all instruments are jointly equal to zero.

The available measure of income is likely to reflect the socioeconomic status of families and this status might be correlated with child health and schooling. At the same time there might be differential responses to the rainfall shock depending on the socioeconomic status of households.

The estimates in Table 2 show that a rainfall shock significantly affects child health in all regions except the Northern Mountain region. Rainfall shocks have positive effects on child health in some regions but negative effects in others.¹⁴ In particular, positive effects are estimated in the Red River Delta and the Northern Central regions while negative effects are estimated in the remaining regions. The coefficients of the rainfall shock effects are jointly significantly different from zero and the F-statistics of 13.92 (with food expenditure) or 11.28 (without food expenditure) exceed the conventional threshold for small finite sample bias (Cameron and Trivedi, 2005, p. 109).

Maccini and Yang (2009) found a positive effect of rainfall on adult height in the Indonesian sample. In the context of Vietnam it is important to take into consideration the regional variation in the composition of agricultural production in order to understand the effect of rainfall shock on child health. In the Red River Delta and North Central regions the most important crop is rice. The estimates in Table 2 show a significant and strong positive effect of rainfall on child health in these two regions. It is likely that in these rice-based regions a shock in rainfall affects health via the income channel. A positive shock is associated with a better rice harvest and possibly better children's nutrition while exposure of the household to drought could be adverse to child health. On the other hand, a very large positive shock such as flooding could cause health environment problems and possibly harm health. In these two regions, the drought effect is dominant resulting in a positive coefficient estimate for these regions. However, this is

¹⁴As noted above, we measure a child's z-score based on age in years. Since we know the birth month and year, we are also able to calculate the z-score based on age in months. Doing so does not significantly affect the estimates of the model. In particular, in the first stage the coefficients of the rainfall shock variables reported in Table 2, column 2 do not significantly differ from the point estimates of the coefficients we obtain when the z-score is based on age measured in months. We calculate $F(5, 129)=1.13$, $Pr.=0.345$. However, when we measure age in months, some of the coefficient estimates of the rainfall shock variables are estimated less precisely. It appears that some of the birth month information is reported with random measurement error.

not the case for the Mekong delta which is the largest rice production area of the country. A positive rainfall shock in this region has an adverse effect on child health. This contradictory result could be explained by recognizing those different ecological characteristics of the Mekong delta as mentioned earlier. It is likely that positive shocks in rainfall cause undesirable floods in the Mekong delta and this could adversely affect health either via a negative income shock or via the deterioration of the health environment.

The estimates in Table 2 indicate that child health is negatively related to a shock in birth year rainfall in the Southern Central, Central Highlands and Southeast regions. Farmers in these regions specialize in long term industrial crops such as coffee, rubber and perennial fruit trees whose productivity may be less sensitive to rainfall variations than is the production of rice. For example, planting coffee trees under taller shade trees may lessen the impact of a drought on the coffee trees (ASEAN Coffee, 2008). As we argued above, it is likely that those cash crop growers are better able to insure against income risks through access to credit relative to small farmers in rice growing regions. Therefore, a negative shock in rainfall is not likely to affect child health through the income channel. On the other hand, it could be that higher rainfall is associated with other adverse health conditions such as increasing incidence of diseases or contaminated drinking water and that those factors negatively affect child health.

In short, we found a significant positive effect of rainfall on child health where farming relies heavily on rain-fed rice production. Therefore, controlling for regional effects helps reconcile our finding with the results in Maccini and Yang

(2009).¹⁵ Maccini and Yang pointed out that Indonesian farmers rely on rainfall to plan their crop activities which is a similar strategy for rain-fed rice farming in Vietnam.

Effect of Child Health on the Delay of School Entry

Table 3 presents estimates of the effect of child health on years of school entry delay. Years of school entry delay are not censored in our sample. In the first two columns, we present the single equation OLS estimates with fixed effects (OLS-FE). These regressions do not control for the endogeneity of child health. As expected, years of entry delay are significantly and negatively associated with child health. A one unit increase in standardized height for age is expected to decrease years of entry delay by

¹⁵ It is possible to compare our estimate to that in Maccini and Yang (2009). Assuming one standard deviation of height at age 5 is equal to 4.6 cm then a shock in rainfall by 20% over the annual mean level induces an increase in child height of 0.41 cm in the region of Red River Delta and North Central ($0.443 \times 0.2 \times 4.6$). In Maccini and Yang (2009), this similar rainfall shock causes an increase in height by 0.57 cm for women in Indonesia.

Table 3 – Estimates of the School Entry Delay Equation

Left hand side variable: Years of Entry Delay

Variables ^a	Single equation with p.c. food expenditure		Single equation without p.c. food expenditure		2SLS with p.c. food expenditure		2SLS without p.c. food expenditure	
	b	se ^b	b	se ^b	b	se ^b	b	se ^b
Gender (1 if male)	0.034	0.02	0.033	0.02	-0.039	0.05	-0.036	0.04
Ethnicity (1 if Kinh)	-0.330***	0.06	-0.331***	0.06	-0.351***	0.07	-0.351***	0.06
Father's schooling	-0.020***	0.01	-0.021***	0.01	-0.015*	0.01	-0.014**	0.01
Mother's schooling	-0.033***	0.01	-0.034***	0.01	-0.036***	0.01	-0.035***	0.01
Number of siblings	0.021	0.01	0.024+	0.01	0.009	0.01	0.005	0.01
Distance to nearest primary school	0.003	0.02	0.004	0.02	-0.012	0.02	-0.012	0.03
Teachers per class	-0.21	0.15	-0.213	0.15	-0.113	0.19	-0.109	0.16
Fraction of poor classrooms	0.022	0.1	0.019	0.1	0.037	0.1	0.041	0.09
Log of per capita food expenditure	-0.059	0.05			0.084	0.09		
Height-for-age zscore	-0.096***	0.02	-0.098***	0.02	-0.523**	0.2	-0.516**	0.16
F	16.78		18.50					
χ^2					136.94		159.93	
p-value	0		0		0		0	
N	4754		4754		4754		4754	

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

^a An overall intercept and fixed effects for region and birth year are also included.

^b The estimates of the standard errors take account of the clustering of observations by commune. Standard error estimates are, in addition, bootstrapped for 2SLS. Thus the estimates are efficient despite the fact that the first and second stages are not estimated jointly.

approx. 0.1 years. Education of parents plays some role in reducing the delay of school entry of their children. The coefficient estimates indicate that a child is likely to save between 0.02 and 0.03 years from avoiding the entry delay, induced by a one year increase in his or her parental schooling attainment. This effect is almost similar to the marginal effect of having one less sibling. Members of minority ethnic groups suffer considerably longer school entry delay relative to members of the majority ethnic group. The magnitude of this effect is almost three times larger than that of health status. The estimates show no difference in years of delayed entry between boys and girls. Schooling conditions do not seem to affect the outcome of years of delayed entry. Distance to nearest primary school, more teachers per class or better classrooms do not have statistically significant effects on years of entry delay. Per capita food expenditure also has no significant effect on school entry delay. This result is not surprising since current family income may be a poor proxy for income at the time when school entry decisions were made.

As noted earlier, the single equation fixed effect approach controls for regional and cohort unobserved factors but can not control for individual heterogeneity. The 5th and 7th columns of Table 3 present the two stage least squares estimates with fixed effects (2SLS-FE) that control for the possible endogeneity of child health^{16,17}. Relative to the model that does not account for the endogeneity of child health, the coefficient estimate of height for age becomes much larger (more negative) and is highly

¹⁶ Controlling for individual heterogeneity by specifying individual fixed effects is not practical because even though we have longitudinal data for a subset of the sample, school entry delay and many regressors do not vary over time and an important outcome equation and the coefficients of many regressors could not be identified with the standard fixed effects methodology.

¹⁷ We conducted Rivers-Vuong tests that indicate that the 2SLS-FE model is the preferred specification relative to OLS-FE. According to this test the coefficient of the residual (s.e.) of the first stage in a schooling delay equation with food expenditures 0.431 (0.21) and 0.421 (0.20) without food expenditure.

significant. This indicates that the effect of child health is attenuated (biased towards zero) in the single equation estimation. Similar to the OLS-FE model, the effect of gender is not significant. The effect of belonging to a minority ethnic group is highly significant again and its magnitude is slightly larger relative to the single equation estimate. Ethnic minority children still delay their schooling entry longer. The estimates of family effects change slightly compared with their OLS-FE counterparts. Parental education still has a negative and significant effect on entry delay. The effect of number of siblings is no longer statistically significant at the conventional level and its size is also reduced. Similar to the OLS-FE model, the effects of school variables and of per capita food expenditure are still insignificant.

In order to illustrate more clearly the effect of child health on years of delayed school entry we present some predictions. Suppose that the average of height for age in the current Vietnamese sample is made equal to that of the reference median height for age. This is equivalent to an increase in the z-score by 1.7 standard deviations with respect to the American child's height for age distribution (see Table 1). In other words, the average height of the 5 year old Vietnamese child population is increased by 7.8 cm (1.7×4.6) under this scenario. The single equation estimate (without food expenditure) shows that these healthier children would be able to close a gap of only 0.17 years of delayed entry (1.7×0.098) out of the current sample mean of 0.41 years which is just about 41% of the total amount of delayed entry. This impact on education seems modest given the substantial increase in height. However, the 2SLS-FE estimate (without food expenditure) implies a much bigger reduction in entry delay of 0.88 year. A more realistic change would be if the z-score increases by 0.62 then the OLS-FE

estimates imply a reduction of the delay by 0.06 (approx. 15 percent of the average delay) while 2SLS-FE imply a reduction by 0.32, (approx. 78 percent of the average delay).

An alternative way to analyze the effect of child health on school entry delay is to see how the probability of starting school without delay is affected by early life health conditions and other variables¹⁸. These estimates of a linear probability regression model of entry delay are presented in Table 4. In the linear probability model inclusion of fixed effects and application of simultaneous equation techniques are straightforward. In the first column, we report estimates of a single equation model with fixed effects including per capita food expenditure. The 5th column of Table 4 presents the corresponding two stage least squares estimates¹⁹. The 3rd and 7th columns report estimates of the models without per capita food expenditure. The estimates of the standard errors in both equations take account of the clustering of the observations at the level of commune²⁰. Improved child health has a negative effect on the probability of school entry delay. The estimated magnitude of the effect of the improvement of early childhood health on schooling enrollment depends on the estimation method. As the estimates in Table 4 show, the 2SLS-FE estimation method provides us with considerably larger estimates in absolute value, by a factor of approximately four, than the estimates obtained by the single equation estimation method. An increase in height at age five by

¹⁸ Alderman et al. (2009) modeled school entry delay as an ordered discrete outcome and estimated an ordered probit model. We found that fixed effects are important and wish to control for them. Ordered probit models with fixed effects pose computational challenges. To sidestep those problems we model separately the outcomes of length of school entry delay and whether or not there is a delay whose estimates are reported in Tables 3 and 4, respectively.

¹⁹ As with years of entry delay equation, we conducted Rivers-Vuong tests that indicate that the 2SLS-FE model is the preferred specification relative to OLS-FE for the probability of delay (Table 4). According to the tests the coefficient of the residual (s.e.) of the first stage in a probability of delay equation with food expenditure is 0.162 (0.07) and 0.157 (0.08) in the equation without food expenditure.

²⁰ The standard error estimates change very little if we control, instead, for heteroscedasticity across individuals (White estimator).

Table 4 – Estimates of the Probability of Entry Delay Equation

Left hand side variable: Delay (1), No Delay (0)

Variables ^a	Single equation with p.c. food expenditure		Single equation without p.c. food expenditure		2SLS with p.c. food expenditure		2SLS without p.c. food expenditure	
	b	se ^b	b	se ^b	b	se ^b	b	se ^b
Gender (1 if male)	0.027*	0.01	0.027*	0.01	0.00	0.02	0.001	0.02
Ethnicity (1 if Kinh)	-0.135***	0.03	-0.136***	0.03	-0.143***	0.04	-0.143***	0.04
Father's schooling	-0.013***	0.00	-0.013***	0.00	-0.011**	0.00	-0.011**	0.00
Mother's schooling	-0.017***	0.00	-0.017***	0.00	-0.018***	0.00	-0.017***	0.00
Number of siblings	0.014+	0.01	0.015*	0.01	0.009	0.01	0.008	0.01
Distance to nearest primary school	0.005	0.01	0.005	0.01	0.00	0.02	0.00	0.01
Teachers per class	-0.079	0.08	-0.081	0.08	-0.043	0.07	-0.042	0.1
Fraction of poor classrooms	0.034	0.05	0.033	0.05	0.04	0.06	0.041	0.05
Log of per capita food expenditure	-0.028	0.03			0.025	0.04		
Height-for-age zscore	-0.048***	0.01	-0.049***	0.01	-0.208**	0.07	-0.205*	0.1
F	17.03		17.88		127.34		219.23	
χ^2					0		0	
p-value	0		0		0		0	
N	4754		4754		4754		4754	

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

^a An overall intercept and fixed effects for region and birth year are also included.

^b The estimates of the standard errors take account of the clustering of observations by commune. Standard error estimates are, in addition, bootstrapped for 2SLS. Thus the estimates are efficient despite the fact that the first and second stages are not estimated jointly.

4.6 cm²¹ reduces the probability of being delayed by almost 5 percent according to the single equation estimates. Using a similar calculation as earlier, we found that the proportion of school entry delay among the 5 year-old child population of Vietnam is reduced by about 8.2 percent (i.e. 0.048×1.7) in a scenario in which there is no disparity in the health status of a Vietnamese average child and that of a reference American child.

Turning to other parameter estimates in Table 4, we found that parental schooling negatively affects the probability of being delayed by 1.3 to 1.7 percent while being a male would increase this probability by 2.7 percent. A family with an additional child would experience an increase in the probability of delay by 1.5 percent. A minority ethnic group member is 14 percent more likely to delay school entry than a majority ethnic group counterpart. The coefficient of per capita food expenditure is not significant.

The 2SLS-FE estimates are of a similar effect of parental schooling on the probability of being delayed as obtained for the single equation estimates. However, the gender coefficient is no longer significant. On the other hand, the probability of delay increases further with minority ethnicity status. The coefficient estimate of the height-for-age z-score is 4 times larger in size meaning that a 4.6 cm increase in height at age 5 reduces the probability of being delayed by 21 percent. In short, we conclude that on average, a relatively healthier child who is a member of the majority ethnic group and whose parents have more education is more likely to start school on time. The estimates also suggest that improved health status during early childhood has a larger impact on the reduction of school entry delay relative to other factors such as ethnicity status or changes in the levels of parental schooling.

²¹ 4.6 cm is one standard deviation of the height of five-year-olds in the reference population.

Effect of Child Health on the Schooling Gap

Estimates of the schooling gap model are presented in Table 5.

As explained earlier, the measure of schooling gap avoids the censoring of the dependent variable. To examine the link between family income and child schooling progress, the first and the third regressions in Table 5 include the variable of per capita food expenditure in logarithmic form.²²

In the first regression of Table 5, which is a fixed effects single equation model, there is no significant difference in the grade gap of boys relative to girls since the gender coefficient estimate is not statistically significant. Children from minority ethnic groups experience a larger grade gap of 0.56 years than Kinh children and this is a statistically significant effect. Parental education has a positive and significant effect on their children's schooling progress. An additional year of schooling that parents complete could help their children close 0.06 to 0.08 years of schooling gap. One less sibling results in a 0.097 year smaller schooling gap. Furthermore, family income unambiguously increases a child's schooling progress. This coefficient estimate suggests that an increase in family income by 100 percent helps close about 0.389 of a year (or 4.58 months) in the gap of child's schooling progress. All the coefficients of school characteristics are not significant. In this regression, the effect of child health is statistically significant. Its estimate suggests an increase of 0.21 years of education as a result of 4.6 cm increase in height at age 5. This estimate thus suggests that achieving the

²² Food expenditure may be a better proxy for the resources available to families at the time they were making decisions on their child's grade gap than for the resources available further back in time when they were making decisions on health care (or early childhood nutrition) and school entry delay.

Table 5 – Estimates of the Schooling Gap Equation

Left hand side variable: years of shortfall of schooling

	Single equation with p.c. food expenditure		Single equation without p.c. food expenditure		2SLS with p.c. food expenditure		2SLS without p.c. food expenditure	
	b	se ^b	b	se ^b	b	se ^b	b	se ^b
Gender (1 if male)	0.028	0.04	0.019	0.04	-0.045	0.07	-0.038	0.07
Ethnicity (1 if Kinh)	-0.563***	0.1	-0.566***	0.1	-0.584***	0.11	-0.583***	0.1
Father's schooling	-0.060***	0.01	-0.066***	0.01	-0.056***	0.01	-0.060***	0.01
Mother's schooling	-0.076***	0.01	-0.080***	0.01	-0.079***	0.01	-0.082***	0.01
Number of siblings	0.097***	0.02	0.118***	0.02	0.085**	0.03	0.102***	0.03
Distance to nearest primary school	-0.003	0.04	-0.002	0.04	-0.018	0.05	-0.015	0.05
Teachers per class	-0.251	0.19	-0.271	0.2	-0.153	0.22	-0.185	0.22
Fraction of poor classrooms (%)	-0.053	0.18	-0.075	0.18	-0.039	0.23	-0.057	0.19
Log of per capita food expenditure	-0.389***	0.09			-0.245+	0.13		
Height-for-age zscore	-0.206***	0.03	-0.222***	0.03	-0.636*	0.31	-0.570+	0.33
F	36.91		39.71					
χ^2					266.19		271.5	
p-value	0		0		0		0	
N	4754		4754		4754		4754	

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

^a An overall intercept and fixed effects for region and birth year are also included.

^b The estimates of the standard errors take account of the clustering of observations by commune. Standard error estimates are, in addition, bootstrapped for 2SLS. Thus the estimates are efficient despite the fact that the first and second stages are not estimated jointly.

same health status as in the reference child population would be expected to earn an average Vietnamese child about 0.35 years of education.

The third regression in Table 5 presents the 2SLS-FE estimates²³. There are no substantial changes in all coefficient estimates except the coefficients of child health and family income. After controlling for endogeneity, the effect of child health becomes considerably larger. A one standard deviation increase in height-for-age (4.6 cm at age 5) corresponding to an increase in the z-score by one unit is expected to reduce the schooling gap by 0.64 years. This reduction in the schooling gap is almost three times larger than estimated with the OLS-FE model. The effect of household income declines both in magnitude and significance level. The 2SLS-FE estimate suggests that a 100 percent increase in family income is predicted to reduce the child's schooling gap by about 0.245 of a year (or 2.94 months). This change indicates that the estimate of the income effect is sensitive to the unobserved heterogeneities reflected in cohort and regional effects as well as their interaction.

We recognize that the presence of food expenditure in the schooling gap equation might cause bias. If income is endogenous, its inclusion is expected to affect the coefficients of the ~~child health~~ **schooling gap** equation. However, if excluding food expenditure as a regressor does not change the estimates of the other parameters very much, this would indicate that simultaneous equations bias is not serious. In the regressions reported in the 3rd and 7th columns of Table 5, we exclude food expenditure from the model, re-estimate it and observe the changes in the estimates. The estimates of the coefficients of the remaining regressors not change greatly as a result of this

²³A Rivers-Vuong test indicates that the simultaneous equation estimates are preferred relative to the single equation estimates of the model with food expenditure. We have a residual coefficient (s.e.) of 0.433 (0.24).

exclusion. This result is also consistent with the finding in Edmonds and Pavcnik (2005) that there is a causal effect of income on child labor but not in the other direction. We conclude, therefore, that simultaneous equation bias arising from the endogeneity of income is not serious.

Having food expenditure in the education equation makes the 2SLS-FE estimates more robust since any possible correlation between the instrumental variables with food expenditure (a proxy for family income) is now accounted for. For this reason, we consider the 2SLS-FE estimate of height for age in the regression with food expenditure as our preferred estimate. Finally, the estimates show that health status still has the most influential effect on schooling progress.

Robustness Checks

We explored whether or not the estimates differ by gender. We found a significant gender shift parameter (intercept) in the child health equation (Table 2) but not in the other equations. Our main interest is in testing whether or not the effect of child health on the two schooling differs by gender. We estimated separate schooling outcome regressions by gender and conducted formal tests of the hypothesis that the coefficients of height-for-age z-score are equal for a given outcome equation across genders. In all cases we cannot reject the null hypothesis that the effects are equal across genders²⁴.

²⁴ We conducted Wald tests that the coefficient of the z-score in each equation is equal to the point estimate of the corresponding coefficient for the other gender we obtain:

	Men	Women
School Entry Delay	$\chi^2(1)=0.00$, Pr=0.953	$\chi^2(1)=0.00$, Pr=0.962
Delay	$\chi^2(1)=0.19$, Pr=0.661	$\chi^2(1)=0.15$, Pr=0.695
Schooling Gap with Income	$\chi^2(1)=2.01$, Pr=0.157	$\chi^2(1)=1.26$, Pr=0.261
Schooling Gap without Income	$\chi^2(1)=1.45$, Pr=0.229	$\chi^2(1)=0.94$, Pr=0.333

Thus our preferred specification is as reported in Tables 2-5. We note that Maccini and Yang (2009) found significant differences in the outcome equations by gender. Specifically the child health affects later life outcomes for women but not for men in Indonesia. Alderman et al. (2009) find smaller child health effects for boys for schooling outcomes for Tanzania, Alderman et al. (2001) find similar differences by gender for Pakistan while Yamauchi (2008) documents gender differences in South Africa. In contrast with these studies we find the same fundamental relationship for both genders. The only gender difference we document is a male disadvantage relative to the reference population: a significant negative intercept in the standardized height (first stage) equation. Our evidence indicates that Vietnamese families treat their sons and daughters equally when it comes to the allocation of health resources while some of the societies studied in the other previous studies favor sons over daughters.

Several previous studies on the effect of child health on schooling outcomes used alternative instruments such as price shocks (Alderman et al. (2009), or Alderman et al. (2001)) the availability of health services (Yamauchi, (2008)) or mother's height (Glewwe and Jacoby (1995)). We tried using these alternative instruments in order to explore the sensitivity of the estimates to the choice of instruments. In order to make sure that prices are exogenous to the outcome being studied we used lagged prices for the subset of the respondents who were also interviewed in the 1992/1993 Vietnam Living Standards Survey. Our sample diminishes substantially and we lose some precision. However, our estimates are immune to the criticism that choices families make now affect commodity prices now which would render the instruments invalid. In all cases

In all cases we cannot reject the null hypothesis that the coefficient for a given equation is equal across genders.

the instruments have insignificant coefficients in the first stage. Thus simultaneous equation estimates based on these weak instruments are likely to be biased (Bound et al. 1995). Mother's height is significant in the first stage but may not be a valid instrument as discussed by Glewwe and Miguel (2008). We also note that the rainfall shock region-specific variables we use as instruments (Table 2) also have significant coefficients in the regression that uses the smaller longitudinal sample that was used to obtain lagged prices: in other words with the same sample, commodity prices are weak instruments while the rainfall shock variables are strong instruments.

We explored some alternative specifications of the rainfall shock variables. One alternative consisted of using the rainfall shock variables (separate coefficients by region) that measured the rainfall shock in the year of birth, but also one year later and one year earlier. Our objective was to investigate whether or not rainfall shocks have a more complex temporal pattern than the model reported in Tables 2-5. The coefficients of the rainfall shock variables for the year of birth changed very little relative to those reported in Table 2. The rainfall shock coefficients for other years were jointly not significantly different from zero²⁵. We conclude that the specification of the rainfall shock variables in Table 2 is robust. As an alternative we used a measure of estimated rainfall at the closest grid point. The results did not materially differ from those we report in Table 2.

Discussion

An important long term goal of development policy has been to lower the incidence of stunting among the population of children five years of age or younger

²⁵ $F(10, 129) = 1.50$, $\Pr = 0.1468$.

(UNICEF 2009). What can be said about the economic benefit from this policy intervention given the estimates of the effect of child health on those two educational outcomes? In the case of Vietnam, the government has set the goal of reducing the stunting rate among children five years old or younger to 20 percent in the year 2010. This target is roughly equivalent to a reduction of the general stunting rate to 30%.²⁶ Inference from the VLSS 1998 sample indicates that this is associated with an increase in the mean z-score of 0.5 units.²⁷ To obtain the return to education, we use an estimate from the World Bank (1996) study on education in Vietnam (Annex 5 – Table 5.1.1). This study suggests a private rate of return to schooling of 4.8%. Based on our estimates in Table 5, an increase in the z-score of 0.5 is predicted to help a child close the schooling gap by 0.32 years which results in an annual wage increase of about 1.54%. In other words, this back of the envelope calculation suggests that early childhood nutrition intervention is likely to have a considerable impact on wages and thus contribute to the long term improvement in living standards.

As reported earlier, Rivers-Vuong tests²⁸ indicate that the models that treat child health as an endogenous variable are preferred to the single equation estimates. These tests also indicate that unobservables that result in better child health are positively correlated with the unobservables that result in a longer school entry delay, higher probability of delay and a larger schooling gap; all worse outcomes. This finding implies that child labor may play an important role as outlined in the framework section above.

²⁶ Calculations with VLSS 1998 show that the under 5 stunting rate is about 28% compared with the sample mean of general stunting rate of 38%.

²⁷ We inferred this number from the assumed normal distribution of height-for-age z-score in the sample.

²⁸ These tests involve estimating the schooling outcome equation with actual z-score as a regressor and the residual of the first-stage regression as an additional regressor. Significance of the coefficient of the residual indicates that simultaneous equation bias would be significant if we failed to control for it, the sign of the coefficient of the residual reflects the sign of the correlation of the error terms of equations (1) and (2).

Healthier children are readier to start school on time and make timely progress but are also more productive as helpers on the family farm. The latter effect appears to predominate, detracting from the schooling outcomes. We leave examination of the relationship between child health and child labor to future work. Of course, measurement error probably also plays a role but it is difficult to explain the substantial differences between the OLS and 2SLS z-score coefficients on the basis of measurement error alone. We expect that measurement error in the data is of moderate importance. Therefore behavioral responses such as the interplay of school readiness and child labor (with the effect of the latter predominating) are likely to play a role in addition to measurement error in explaining the substantial differences in the single equation and simultaneous equation estimates of the effect of child health on schooling outcomes (α).

It is informative to compare our estimates to those in two other recent studies that used data from other developing countries. The study of Alderman et al. (2009) differs from ours in that it employs an ordered probit model and uses the percentage of the reference median height in place of standardized height for age in the regressions. However, these authors pointed out that stunting in their sample occurred at the level of 85% of median height and they predicted a drop in 0.74 years of delay associated with an increase up to 95% of median height. This amounts to approximately 0.8 years if the population height in Tanzania is predicted to be the same as the median height and therefore equivalent to an increase in 2 units of standardized height-for-age.²⁹ Our estimates for Vietnam show that such an increase of the z-score leads to about 1.05 year reduction in delay (0.523×2). Therefore, our estimate of the effect of child health on school entry delay is somewhat larger than theirs.

²⁹ Recall that we defined stunting as having a z-score below -2.

Alderman et al. (2001) estimated a probit model for the probability of delay. Their marginal effect estimate for the preferred estimate of child health effect shows that an increase in the z-score by 0.5 results in a decrease in the probability of delay of about 2 percent for boys and about 9 percent for girls in Pakistan. Our 2SLS-FE estimate using the linear probability model indicates about 10.4 percent (0.208×0.5) decrease in the probability of delay in response to an increase in the z-score by one-half unit. Therefore, the impact of child health is estimated to be larger in our model.

Conclusion

We estimated models of the relationship between child health and schooling outcomes with data from Vietnam. We treated child health as an endogenous variable and established that simultaneous equation estimates of the effect of child health on schooling outcomes are much larger than single equation estimates. We used rainfall shocks as an exogenous instrument to identify our equations and found that rainfall has a heterogeneous impact on child health in different regions of Vietnam in ways that reflect regional economic differences. We studied schooling outcomes that measure school entry delay and the schooling gap (or shortfall in grade progress for age). Better child health results in a smaller school entry delay, lower probability of delay and better progress through the grades. The predicted magnitudes are both statistically and empirically significant and imply that a policy intervention that would improve child health to a realistic degree would have a considerable impact on wages. We do not find significant gender differences in the effect of child health in the outcome equations. We also find that members of minority ethnic groups suffer delayed school entry and slower progress through school.

References

- Alderman, Harold, Jere R. Behrman, Victor Lavy, and Rekha Menon (2001). "Child health and school enrollment," *Journal of Human Resources*, 36(1), 185-205.
- Alderman, Harold, John Hoddinott, and Bill Kinsey. (2006). "Long term consequences of early childhood malnutrition," *Oxford Economic Papers*, 58(3), 450-474.
- Alderman, Harold, Hans Hoogeveen, and Mariacristina Rossi (2009). "Preschool Nutrition and Subsequent Schooling Attainment: Longitudinal Evidence from Tanzania," *Economic Development and Cultural Change*, 57(2), 239-260.
- ASEAN Coffee (2008).
http://www.coffeeasean.org/details.asp?Object=7&news_ID=191273691.
- Bound, John, David A. Jaeger, and Regina M. Baker. (1995). "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90, (430), 443-50.
- Cameron, A. Colin and Pravin K. Trivedi. (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.
- Edmonds Eric V., and Nina Pavcnik. (2005). "The Effect of Trade Liberalization on Child Labor," *Journal of International Economics*, 65(2), 401-419.
- Glewwe, Paul. (2004). "An Investigation of the Determinants of School Progress and Academic Achievement in Vietnam," *Economic Growth, Poverty, and Household Welfare in Vietnam*, chapter 13, 467-501. The World Bank's Regional and Sectoral Studies No. 29086.
- Glewwe, Paul, and Hanan G. Jacoby. (1995). "An economic analysis of delayed primary school enrollment in a low income country: The role of early childhood nutrition," *Review of Economics and Statistics*, 77(1), 156-169.
- Glewwe, Paul, Hanan G. Jacoby, and Elizabeth M. King. (2001). "Early childhood nutrition and academic achievement: A longitudinal analysis," *Journal of Public Economics*, 81(3), 345-368.
- Glewwe, Paul, and Edward A. Miguel. (2008). "The Impact of Child Health and Nutrition on Education in Less Developed Countries,". In: Schultz, T. Paul, and John Strauss, (Eds.), *Handbook of Development Economics*, vol. 4, Chapter 10, Elsevier/North-Holland, Amsterdam.

Grosh, Margaret and Paul Glewwe. 2000. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Washington, DC: The World Bank.

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

Legates, David R., and Cort J. Willmott. (1990). "Mean Seasonal and Spatial Variability in Gauge-Corrected, Global Precipitation," *International Journal of Climatology*, 10(2), 111-127.

Maccini, Sharon, and Dean Yang. (2009). "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *The American Economic Review*, 99(3), 1006-1026.

Ravallion, Martin and Dominique Van de Walle, (2008). *Land in Transition: reform in rural Vietnam*, The World Bank, Washington, DC.

Rivers, Douglas, and Quang H. Vuong. (1988). "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39(3), 347-366.

Rosenzweig, Mark, R. 1990. "Population Growth and Human Capital Investments: Theory and Evidence." *Journal of Political Economy*, 98(5, Part 2): S38-S70.

Strauss, John, and Duncan Thomas. (1998). "Health, Nutrition, and Economic Development," *Journal of Economic Literature*, 36(2), 766-817.

Strauss, John and Duncan Thomas. (2008). "Health over the Life Course," In: Schultz, T. Paul, and John Strauss, (Eds.), *Handbook of Development Economics*, vol. 4, Chapter 8, Elsevier/North-Holland, Amsterdam.

The World Bank. (1996). "Annex 5.1. Rates of Return for Wage Workers Based on Estimations of the Human Capital Earnings Function". In: *Vietnam Education Financing Sector Study*. Report No. 15925-VN, Human Resources Operations Division, Country Department, East Asia and Pacific Region.

The World Bank. (2001). *Vietnam Living Standards Survey (VLSS), 1997-98 Basic Information*, Poverty and Human Resources Division, Washington, DC: The World Bank.

Trenberth, Kevin E., (1997). "The Definition of El Niño," *Bulletin of the American Meteorological Society*, 78(12), 2771-2777.

United Nations (2000). "Lessons from the 16-Country UNEP Study". In: *Reducing the Impact of Environmental Emergencies through Early Warning and Preparedness: The Case of the 1997-98 El Niño – A UNEP/NCAR/UNU/WMO/ISDR Assessment*.

United Nations Children's Fund (2009). "The Challenge of Undernutrition". In: *Tracking Progress on Child and Maternal Nutrition - a survival and development priority*.

World Health Organization (1999). "Health Impact". In: *El Niño and Health* – WHO/SDE/PHE99.4

Wiens, Thomas B. (1998). "Agriculture and Rural Poverty in Vietnam". In: Dollar, David, Paul Glewwe and Jennie Litvack (eds.) *Household Welfare and Vietnam's Transition*, The World Bank, Washington, DC, 61-98.

Yamauchi, Futoshi. (2008). "Early Childhood Nutrition, Schooling, and Sibling Inequality in a Dynamic Context: Evidence from South Africa," *Economic Development and Cultural Change*, 56(3), 657-682.